

Statistical Methods

*Concepts, Applications
and Computation*

Second Revised Edition

Y.P. Aggarwal

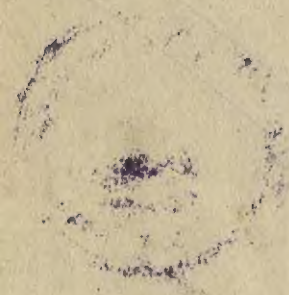
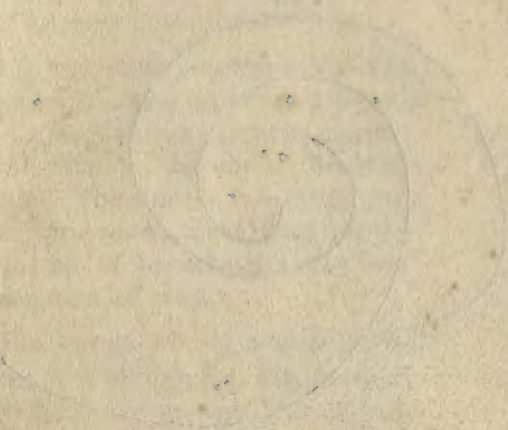
The book presents a simple treatment of statistical methods as applied to education, psychology, sociology, economics, commerce, management, social work and other social sciences. The main emphasis is on generating an understanding of concepts, application and procedures involved in statistical analysis of experimental and observational data. It is the result of the author's long experience of teaching post-graduate classes.

All important parametric and non-parametric tests have been presented in a manner that would be convenient even to the non-mathematical mind. The student can easily grasp the procedures right from the setting up of hypotheses to the interpretation of results.

The second edition is an improved one with all printing and computational errors removed. The diagrams have also been improved.

STATISTICAL METHODS
Concepts, Application and Computation





STATISTICAL METHODS

Concepts, Application and Computation

Y.P. AGGARWAL

M.A., M.Ed., Ph.D. (Ottawa)

Professor of Education

Kurukshetra University



STERLING PUBLISHERS PRIVATE LIMITED

STERLING PUBLISHERS PRIVATE LIMITED

L-10 Green Park Extension, New Delhi-110016

G-2 Cunningham Apartments,

Cunningham Road, Bangalore-500052

Date 21-3-89
Page No. 4323

311
A662

STATISTICAL METHODS: Concepts, Application and Computation

© 1988, Y.P. Aggarwal

Second Edition (Revised and Enlarged)

PRINTED IN INDIA

Published by S.K. Ghai, Managing Director, Sterling Publishers Pvt. Ltd.,
L-10, Green Park Extension, New Delhi-110016.
Printed at Roopak Printers, Delhi-110032

PREFACE TO THE SECOND EDITION

The first edition of the book met with an excellent reception by the students and teachers of various social sciences and was sold out within a year of its publication. In the second edition no attempt has been made to rewrite the book and the original appeal to every conceivable user has been kept intact. The printing and other computational errors have been corrected and the figures and diagrams improved. However, a brief chapter on Second Generation of multivariate analysis has been added.

Suggestions by teachers and students in the field of statistics for the improvement of this edition will be gratefully acknowledged and incorporated in the next edition.

Y.P. Aggarwal

PREFACE TO THE FIRST EDITION

The main objective of this book is to introduce students and research workers in psychology, education, sociology and many other social sciences to concepts, application and computational procedures involved in statistical analysis of experimental and observational data. Students of experimental medicine, psychiatry, biological sciences and to some extent of physical sciences will also find the book useful. An attempt has been made to introduce the student to the practical technology of statistics in a near non-mathematical manner. The quantum of mathematical skill required is a knowledge of the high school algebra.

This book has been designed as a text for both one-semester and full-year courses on statistics. The author has a strong belief that the students of research must be introduced to the techniques of analysis of variance and covariance. These statistical devices find the largest application in experimental research. A chapter on non-parametric statistical techniques has also been included in which some important and more popular non-parametric tests have been presented.

The emphasis of the book is on generating a conceptual, procedural and computational understanding of the statistical techniques right from the setting up of the statistical hypothesis in symbolic and verbal terms to the interpretation of the results. At several places the complete solutions of problems have been shown in boxes to aid understanding through a compact presentation which has been kept at a fairly simple level.

The mind is fed from various streams whose sources may sometimes be not known. However, the contribution of the great statisticians like Fisher, Spearman, Pearson, Yates, Gossett and Wilcoxon is gratefully acknowledged. Several ideas have been borrowed from prestigious texts written by J.P. Guilford, H.E. Garrett, G.A. Ferguson, E.F. Lindquist, S. Siegel, R.B. McCall and A.L. Edwards for whom the author expresses his appreciation and gratitude. My teacher Dr. N. Vaidya and my colleagues especially Dr. C. L. Kundu and Dr. H.C. Sinha have been a great source of inspiration. My thanks are also due to my students and other colleagues, like Dr. S.M. Gupta and Dr. Vijay Kumar who provided solutions of numerical problems on a few chapters. I must acknowledge the help rendered by Sh. D.N. Sharma for the drawings, and Sh. R.G. Gupta for typing the manuscript.

Y.P. Aggarwal

LIST OF TABLES

<i>Table</i>	<i>Description</i>	<i>P. No.</i>
1.1	Characteristics and Examples of Measurement Scales at a Glance.	7
2.1	The Development of a Frequency Distribution.	11
2.2	Relative Frequency Distribution based on Data of Table 2.1 (C).	12
2.3	One Hundred Hypothetical Scores.	13
2.4	Frequency Distribution of 100 Hypothetical Scores with Class Interval (C.I.) size = 10.	14
2.5	Exact Limits and Mid points of Class Intervals.	18
2.6	Cumulative Frequency and Cumulative Percentage Frequency in a Frequency Distribution.	23
2.7	Smoothed Frequencies.	28
3.1	Calculation of Mean from Frequency Distribution with Class Interval of size 1.	36
3.2	Calculation of Mean from Frequency Distribution with Class Interval size of two or more.	37
3.3	Calculation of Mean from a Frequency Distribution by Short Method or Assumed Mean Method.	38
3.4	Deviations from the Mean.	42
3.5	Effect of a Constant on Mean.	43
3.6	Calculation of Combined Mean.	44
3.7	Calculation of Median.	48

<i>Table</i>	<i>Description</i>	<i>P. No.</i>
3.8	Computation of the Median from Distribution with Gaps.	51
4.1	Three Sets of Scores with Equal Means but Different Dispersions.	59
4.2	Calculation of Range.	60
4.3	Calculation of Average Deviation (AD) from Set 3 of Table 4.1	62
4.4	Calculation of SD from Ungrouped Scores.	65
4.5	Calculation of SD by Long Method (Using Real Mean).	66
4.6	Calculation of SD by Short Method (Deviations taken from Assumed Mean).	68
4.7	Calculation of Q_1 , Q_3 and Quartile Deviation.	71
4.8	Data Illustrating Sum of Squares (Σx^2), Variance (V) and Standard Deviation (SD).	75
5.1	Main Types of Norms for Educational and Psychological Tests.	78
5.2	Computation of Percentile Points from Ungrouped Data (Scores of 40 students on an Arithmetic test).	81
5.3	Calculation of P_{10} , P_{30} , P_{40} , P_{50} , P_{60} , P_{70} , P_{80} and P_{90} .	87
5.3A	Computation of Standard Scores.	94
5.4	Computation of Standard Scores with $M=50$, $SD=10$.	95
5.5	The Stanine Score System.	96
5.6	Computation of T Scores.	99
6.1	List of 36 Possible Outcomes when two Dice are thrown.	105
6.2	Calculation of Probability in Different Situations.	106
6.4	The Binomial Coefficients of $(p+q)^n$ Pascal's Triangle.	111

<i>Table</i>	<i>Description</i>	<i>P. No.</i>
6.5	The Binomial Distribution $(p+q)^n$ and $N(p+q)^n$ with $p=.5$, $n=10$ and $N=1024$.	112
6.6	Normal Probability Curve Area Values for given z Values.	118
6.7	Area under Normal Probability Curve between given Limits.	118
6.8	Calculation of Per cent Area and Number of Cases in each Sub-group.	129
7.1	Paired Scores for Three Levels of Correlation.	142
7.2	Calculation of Product Moment r by two different Formulas:	144
7.3	Rank-Difference Coefficient of Correlation (Case of no ties).	147
7.4	Rank-Difference Coefficient of Correlation.	150
7.5	Values of Coefficient of Determination, and Coefficient of Alienation for some Selected Values of r .	152
7.6	Worksheet for the Calculation of r_{bts}	156
7.7	Worksheet for the Calculation of Point Biserial Correlation, r_{pbts} .	158
7.8	Worksheet for the Calculation of Tetrachoric Correlation.	160
8.1	Meaning of Levels of Confidence.	174
9.1	Summary of the Test of Difference of Means of Independent Groups (Arithmetic Ability Example).	187
9.2	Summary of the Test of Difference between Means for Correlated Groups (Attitude Test example : Small Sample).	188
9.3	Summary of the Test of the Difference between Means for Correlated Large Samples Single Group Method.	189

<i>Table</i>	<i>Description</i>	<i>P. No.</i>
9.4	Summary of the Test of the Difference between Means for Correlated Groups (Non Sense Syllable Test example), Difference Method.	191
10.1	Computation of Chi-Square Test of Hypothesis of Equality (Example about atom bomb).	207
10.2	Ratings of two Groups on Leadership Qualities.	219
10.3	Computation of a Median Test for two Samples	224
10.4	The Kolmogrov-Smirnov Test of Similarity of Distributions (Two independent Small Samples in Equal n's)	228
10.5	Kolmogrov-Smirnov Two-Sample Test with Large and Unequal n's.	230
10.6	A General Guide for the Selection of Non-Parametric Tests.	234
11.1	Work-Sheet for One-Way Analysis of Variance on Hypothetical Scores using Deviation Score Method.	243
11.2	Summary of Analysis of Variance.	247
11.2A	Worksheet for One-way ANOVA on Hypothetical Scores (RAW SCORE METHOD).	248
11.3	Worksheet for Two-way Analysis of Variance on Hypothetical Data (RAW SCORE METHOD).	253
11.4	Interaction of Method and Teacher (Hypothetical Mean Achievement Scores).	261
11.5	Subtraction of the Main Effects	263
12.1	Worksheet for Covariance Analysis.	273
12.2	Summary of ANOVA	274
12.3	Summary of ANCOVA	275
12.4	Adjusted γ Means	275

LIST OF FIGURES

<i>Figure</i>	<i>Description</i>	<i>P. No.</i>
2.1	Illustration of Size, Lower and Upper exact Limits and Midpoints of the Interval 100-109	19
2.2	The Coordinate System	21
2.3	Histogram of the 50 Scores given in Table 2.6	24
2.4	A Histogram with Sides of Rectangles not Projected to the Baseline (Data in Table 2.6).	25
2.5	A Frequency Polygon for the Data in Table 2.6	26
2.6	A Frequency Polygon Constructed from a Histogram given in Figure 2.4	27
2.7	Original and Smoothed Frequency Polygons based on Data in Table 2.7.	29
2.8	A Cumulative Frequency Curve based on the Data in Table 2.6.	31
2.9	A Cumulative Percentage Curve or Ogive based on the Frequency Distribution in Table 2.6.	32
3.1	Mean as Centre of Gravity of a Frequency Distribution.	41
3.2	Computation of the Median when there is Duplication of Scores (Even number).	46
3.3	Computation of Median when there is Duplication of Scores (Odd Number).	47
3.4	The Relative Position of Mean, Median and Mode in Different Types of Distributions : (A) Symmetrical Unimodal; (B) Symmetrical Bimodal;	55

<i>Figure</i>	<i>Description</i>	<i>P. No.</i>
	(C) Positively Skewed, and (D) Negatively Skewed.	55
4.1	Comparison of Standard Deviation and Variance.	73
5.1	Determination of Percentile Rank Corresponding to the Score Value of 63.	85
5.2	Determination of Percentile Rank Corresponding to a Score Value of 52.	86
5.3	Interpolation for the Calculation of PR	90
5.4	Cumulative Percentage Curve for the Calculation of Percentiles and PR's.	91
5.5	Stemme Scale showing Standard Deviation Intervals and Percents in each Score from 1 to 9	97
5.6	Various Types of Standard Score Scales in Relation to Percentiles and the Normal Curve	98
6.1	Different Proportions of Area under the Normal Curve	114
6.2	Calculation of Area when z Limits fall on both Sides of the Mean.	120
6.3	Calculation of Area when z Limits fall on one side of the Mean	121
6.4	Calculation of Area above a given Score	122
6.5	Calculation of Area below a given Score	123
6.6	Score Limits Equivalent to Middle 60% Cases	124
6.7	Percentage of Cases Exceeding a Particular Score	125
6.8	Comparison of Relative Difficulty Value of Test Items Based on Sigma Differences	126
6.9	Classification of a Group into sub-Groups	128
6.10	(A) Negative Skewness to the Left (B) Positive Skewness to the Right	131
6.11	(A) Leptokurtic (B) Normal or Mesokurtic, (C) Platykurtic Curves	133
7.1	Scatter Plots for Three Levels of Correlation	142

<i>Figure</i>	<i>Description</i>	<i>P. No.</i>
8.1	Sampling Distribution of Means — Size Variability of Obtained Means and Population M in Terms of σ_M	150
8.2	Distribution of t-Ratios (t-Ratio Distribution) Ranging from 1 to 8.	152
8.3	Confidence Intervals for the Mean of that Distribu- tion with $df = 15$	153
9.1	Two-Sample Distribution of Means — (A) the Critical Region for the Critical Probability (A) Two-tailed or Non-directional Test (B) One-tailed or Directional Test	161
10.1	Distribution of Chi-Square for Different Degrees of Freedom	164
11.1	Graphical Representation of the Frequency based on the Data of Table 11.4	262



CONTENTS

<i>Preface</i>	v
<i>List of Tables</i>	xvii
<i>List of Figures</i>	xxi

CHAPTER 1 : THE STUDY OF STATISTICS

1.1	Statistics	1
1.2	Importance of the Study of Statistics	2
1.3	Parameters and Estimates	3
1.4	Descriptive and Inferential Statistics	3
1.5	Variables and Their Types	4
1.6	Measurement Scales	4
1.6.1	Nominal or Classifactory Scale	5
1.6.2	Ordinal or Ranking Scale	6
1.6.3	Interval Scale	6
1.6.4	Ratio Scale	8
	Exercises for Practice	8

CHAPTER 2 : FREQUENCY DISTRIBUTIONS AND THEIR GRAPHIC REPRESENTATION

2.1	Frequency Distributions	10
2.2	Relative Frequency Distribution	12
2.3	Steps	14
2.4	Exact Limits and Mid-Points of the Class Intervals	17
2.5	Assumptions regarding values within the Intervals	20

Graphic Representation of Data	20
2.6.1 Histograms	22
2.6.2 Frequency Polygons	24
2.6.3 Smooth Frequency Polygon	28
2.6.4 Cumulative Frequency Curve	29
2.6.5 Cumulative Percentage Curve or Ogive	31
Exercises for Practice.	33

CHAPTER 3: MEASURES OF CENTRAL TENDENCY

3.1 The Mean (M)	34
3.1.1 Calculation of Mean by Long Method	36
3.1.2 Calculation of Mean by Short Method or Assumed Mean Method	38
3.1.3 Some Properties of Mean	41
3.2 The Median (M_d)	45
3.2.1 Ungrouped Data	45
3.2.2 Calculation of Median from a Frequency Distribution	48
3.2.3 Calculation of Median when the Frequency Distribution Contains Gaps	50
3.3 The Mode (M_o)	52
3.3.1 Calculation of Mode in a Frequency Distribution	53
3.4 Comparison of the Mean, Median and Mode	54
3.5 Guidelines for the Use of Various Measures of Central Tendency	56
Exercises for Practice.	57

CHAPTER 4: MEASURES OF VARIABILITY

4.1 The Range	
4.2 The Average Deviation (A_d)	61
4.3 The Variance and Standard Deviation	63

4.3.1 Methods of Calculating Variance and Standard Deviation from Ungrouped Data	64
4.3.2 Calculation of SD from the Grouped Data	66
4.3.3 Properties and Uses of Variance and SD as Measures of Variability.	69
4.4 The Semi-Inter-Quartile Range or Q	70
4.5 Relationship Between Sum of Squares, Variance and SD	73
Exercises for Practice	75

CHAPTER 5 : MEASURES OF RELATIVE STANDING

5.1 Age Norms	79
5.2 Grade Norms	79
5.3 Percentiles	80
5.3.1 Calculation of Percentiles from Ungrouped Data	81
5.3.1.1 When no duplication near Percentile exists.	81
5.3.1.2 When Duplication near Percentile exists	83
5.4 Calculation of Percentile Ranks From Ungrouped Data	84
5.4.1 When no Duplication near Percentile exists	84
5.4.2 When Duplication near Percentile exists	85
5.5 Calculation of Percentiles From The Grouped Data or Frequency Distribution	86
5.6 Calculation of Percentile Ranks From Grouped Data	89
5.7. The Cumulative Percentage Curve or Ogive	91
5.7.1 Percentiles and Percentile Ranks from Ogive	92
5.8 Standard Scores	93
5.9 The Stanine	95
5.10 The T Scale	96
Exercises for Practice	100

CHAPTER 6 : PROBABILITY, BINOMIAL DISTRIBUTION AND NORMAL DISTRIBUTION

6.1	Some Fundamental Notions	104
6.1.1	Possible Outcomes	104
6.1.2	Addition and Multiplication Rules	106
6.1.3	Permutations and Combinations	107
6.2	The Binomial Distribution	108
6.3	The Normal Distribution	113
6.3.1	Properties of the Normal Curve	114
6.3.2	The Equation for the Normal Distribution Curve	116
6.3.3	The Unit Normal Curve	116
6.3.4	Areas under the Normal Curve	117
6.3.5	Problems and Numericals on Normal Distribution.	119
6.3.5.1	Cases within given Score Limits	119
6.3.5.2	Limits of Scores which include a given Percentage	124
6.3.5.3	Comparison of two Distributions in Terms of 'Overlapping'	125
6.3.5.4	Determination of Relative Difficulty of Test Items	126
6.3.5.5	Division of a Group into Sub-groups	127
6.3.6	Importance of the Normal Distribution	130
6.4	Divergence From Normality	131
6.4.1	Skewness	131
6.4.2	Kurtosis	132
6.5	Measures of Skewness and Kurtosis Based on Moments Methods	133
6.5.1	Significance of the Measures of Skewness and Kurtosis.	136
6.6	Importance of Measures of Skewness and Kurtosis	136
	Exercises for Practice	

CHAPTER 7 : CORRELATIONAL TECHNIQUES

7.1	The Concept	141
7.2	The Production Moment Correlation, r	143
7.3	Some other Formulas	146
7.4	Spearman's Rank-Order Correlation Coefficient (ρ)	147
7.4.1	Calculation of ρ when no Ties exist	147
7.4.2	Calculation of ρ when tied Ranks exist	149
7.5	Properties of the Correlation Coefficient	151
7.5.1	The Range of r	151
7.5.2	The Coefficient of Determination, r^2	151
7.5.3	The Effect of Origin and Unit upon Correlation Coefficient	152
7.5.4	Correlation and Causation	153
7.5.5	Factors Influencing the Size of the Correlation Coefficient	153
7.5.6	Assumptions underlying the Product Moment Correlation	154
7.5.7	The Interpretation of r in Terms of Verbal Description	155
7.6	Biserial Correlation	155
7.7	Point Biserial Correlation	157
7.8	Tetrachoric Correlation	159
7.9	The PHI Coefficient (ϕ)	162
	Exercises for Practice	163

CHAPTER 8 : SIGNIFICANCE OF MEAN AND OTHER STATISTICS

8.1	Sampling Distribution and the Standard Error of the Mean	166
8.2	Computation of the Standard Error of Mean, SE_M	167

8.3	Application and Interpretation of SE_M in Large Samples	169
8.4	The Distribution of t	171
8.5	Degrees of Freedoms, df .	172
8.6	Levels of Significance	174
8.7	Application and Interpretation of SE_M in Small Samples	175
8.8	The Standard Error of a Median, σ_{Med}	176
8.9	The Standard Error of a Standard Deviation, SE_σ	178
8.10	The Standard Error of Percentages and Proportions	178
8.11	The Standard Error of a Correlation Coefficient	179
8.12	Conversion of r 's Into Fisher's z Function	180
	Exercises for Practice	181

CHAPTER 9: THE SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS AND OTHER STATISTICS

9.1	The Null Hypothesis, H_0	183
9.2	The Process	184
9.3	Standard Error (SE) of the Difference Between Two Independent Means (Large Samples)	186
9.4	The SE of Difference Between Means in Small Independent Sample	187
9.5	Standard Error of the Difference Between Two Correlated Means	189
9.6	Difference Method (Small Sample)	190
9.7	The Significance of Difference Between Standard Deviations	192
9.8	The Significance of the Difference Between Two Independent Proportions	194
9.9	The Significance of the Difference Between Two Correlated Proportions	195

9.10	The Significance of the Difference Between Two r's	197
9.11	Two Tailed and one Tailed Tests of Significance	199
9.12	Type I and Type II Errors	200
	Exercises for Practice	201

CHAPTER 10 THE CHI-SQUARE TESTS AND OTHER NON-PARAMETRIC METHODS

10.1	Degrees of Freedom, df	205
10.2	Test of the Hypothesis of Equal Probability	206
10.3	Test of Hypothesis of Independence (Difference)	208
10.4	Test of the Hypothesis of Normality	209
10.5	Calculation of Chi-Square for 2×2 Tables	211
10.6	Yates' Correction for Continuity	213
10.7	Chi-Square from Percentages	214
10.8	General Observations on Chi-Square	215
10.8.1	Assumptions of the Chi-Square Test	215
10.8.2	One-tailed and Two-tailed situations	216
10.8.3	Reduction of an $R \times C$ Table to a 2×2 Table	216
10.8.4	Additivity of Chi-Square	216
10.9	Non-parametric Statistical Tests	217
10.9.1	Sign Test	218
10.9.2	Sign Test with Large Samples	221
10.9.3	The Median Test	223
10.9.4	A General Non-Parametric Test for two Independent Samples	225
10.9.5	The Kolmogorov-Smirnov Two Sample Test	227
10.9.6	The K-S. Test with Large Samples	229
10.9.7	Some Precautions	...

10.9.8	A Guide for the Selection of Non-Parametric Tests	233
	Exercises for Practice	236

CHAPTER 11 : THE ANALYSIS OF VARIANCE, ANOVA

11.1	The Rationale	240
11.2	One-way or Single Classification Anova	241
11.2.1	Deviation Score Method	243
11.2.2	Raw Score Method	248
11.3	Post-Anova Test of Differences By Use of 't'	250
11.4	Two-Way or Double Classification Anova	252
11.4.1	Effect of Introduction of a Second Factor	256
11.5	Notation For Three-Way Anova	257
11.6	Interaction	260
11.7	Assumptions Underlying the Analysis of Variance	264
11.8	General Uses and Limitations of Anova	265
	Exercises for Practice	266

CHAPTER 12 : THE ANALYSIS OF COVARIANCE

12.1	Introduction	270
12.2	Computation	272
12.3	Notation and Description of Computational Steps	276
12.4	Assumptions Underlying Ancova	281
12.5	General Uses of Ancova	281
	Exercises for Practice	283

CHAPTER 13 : RELIABILITY AND VALIDITY OF TEST SCORES

13.1	Reliability	284
13.1.1	Methods of Estimating Reliability,	284

13.1.1.1	Test-Retest Method	285
13.1.1.2	Alternate or Parallel form Method	285
13.1.1.3	The Split-half Method	286
13.1.1.4	'Rational Equivalence' Method	287
13.1.2	Factors Affecting Reliability	289
13.1.2.1	Length of test	289
13.1.2.2	Range of Talent	290
13.1.2.3	Testing Conditions	290
13.2	Validity	290
13.2.1	Types of Validity	291
13.2.1.1	Content Validity	291
13.2.1.2	Face Validity	291
13.2.1.3	Concurrent Validity	291
13.2.1.4	Criterion Related Validity	292
13.2.1.5	Construct Validity	293
13.2.1.6	Factorial Validity	294
13.2.2	Factors Affecting Validity	294
13.3	Relation Between Reliability and Validity	294
13.4	Item Analysis	295
13.4.1	Item Difficulty	295
13.4.2	Item Discrimination	295
	Question and Practice	298

CHAPTER 14 REGRESSION AND PREDICTION

14.1	History and Meaning	300
14.2	Equation of a Straight line	301
14.3	Simple Regression	304
14.4	Regression Equations from Raw Scores	304
14.5	Regression Equations from SD's, r and M 's	307
14.6	Relationship between b coefficients and r	309
14.7	Standard Error of Estimate	310
14.8	Assumptions	311
14.9	Multiple Prediction	311

14.10	The coefficient of Multiple Correlation, R	312
14.11	The Multiple Regression Equation	314
14.12	Calculation of R from Betas	316
14.13	Standard Error of Estimate for Multiple Prediction	316
14.14	Other Methods	317
	Exercise for Practice	317
15.1	Distinguishing Characteristics	319
15.2	Second Generation Methods: An obvious extension of First Generation Techniques	320
15.3	Issues of Variables and their relationships in the Second Generation Methods	320
15.4	Concluding Remarks	321
	<i>Appendices</i>	322
	<i>Answer to Exercises for Practice</i>	365
	<i>Bibliography</i>	372
	<i>Index</i>	375

CHAPTER 1

THE STUDY OF STATISTICS

1.1 Statistics

The word 'Statistics' appears to have been derived from the Latin 'Status' meaning a '(political) state'. In its origin, therefore, statistics was simply the collection of numerical data, by the kings, on different aspects useful to the state. With the passage of time, however, its scope began to include collection of numerical data pertaining to almost every endeavour, calculations of percentages, etc. and the presentation of data in tables and charts. We could then hear about statistics of births and deaths, imports and exports, of marriages and divorces, of mental abilities in a person, of population and the like. By the end of the 19th century, statistics began to concern itself not only with the collection and presentation of data but also with interpretation and drawing of inferences from the data.

Today 'Statistics' is the scientific study of handling quantitative information. It embodies a methodology of collection, classification, description and interpretation of data obtained through the conduct of surveys and experiments. The essential purpose is to describe and draw inferences about the numerical properties of populations. The term population is defined in a more general and broader sense and includes not only the common place meaning as groups or aggregates of people or living things but also groups or aggregates of trees, animals, soil, birds, responses to test items, books, buildings and the like. The populations can be infinite when enumeration or listing up of all the elements is impossible or extremely difficult, for example fish in the sea, stars in the sky, or trees in the forest.

Finite populations have enumerable elements such as students in a school and cards in a deck.

Statistics is concerned with the quantifiable properties of populations, that is, the properties to which numerals can in some manner be assigned. Populations are generally very large in size leading to the impossibility of producing numerical estimates or statistics based on all elements or members. Study of a complete population may be too expensive, time-consuming and full of hazards of inaccuracy. Hence, the statistician draws a sub-group, or subaggregate or portion of the population by using some appropriate method. It is called a 'sample'. He studies the sample and proceeds to generalize the results over to the whole population from which the sample was drawn. It may involve some marginal error, the magnitude of which can be estimated by appropriate numerical procedures.

1.2 Importance of the Study of Statistics

Statistical thinking and operations in behavioural sciences research are important from a variety of standpoints. Statistics permit the use of a descriptive language which is more efficient and exact in communication. They disallow any vague conclusions and emphasize arriving at definite ones. These techniques enable us to present our results in a summarized, more meaningful and convenient form and thus bring order out of chaos. Statistics further enable us to draw generalizations and make predictions. Complex and bewildering events can be analysed and causal factors identified. Leaving aside a few solitary fields, research in behavioural sciences will be poorer without the use of statistical analysis.

The study of statistics is gaining further impetus because of the following facts :

1. Professional literature is replete with statistical symbols, concepts and ideas.
2. All advanced courses of study require a formal course work in statistics.
3. The professional training of a behavioural scientists includes the requirements of the study of statistics.

4. Statistics are widely used in research in all fields of human knowledge. They are making further in-roads into fields not so far covered by them.

1.3 Parameters and Estimates

Measurements of samples generate some numerical values like an arithmetical average. These values are termed as *estimates* or *statistics*. Values which are descriptive of the populations are called *parameters*. Parameters are generally estimated from sample statistics but the former do exist. These may, however, remain unknown for reasons of convenience. The distinction between parameter and statistics or estimate reflects itself in statistical notion. Generally, statisticians use Greek letters to represent parameters and Roman letters to represent estimates or statistics. For example the symbols σ and μ (Greek letters Sigma and Mu) may be used to represent population standard deviation and mean respectively. Correspondingly, the symbols S and M are employed to represent estimates or statistics based on samples for the two measures. However, variations in this usage may be encountered in the statistical literature.

1.4 Descriptive and Inferential Statistics

Descriptive statistics refers to procedures for organizing, summarizing and describing quantitative data about the samples or about the populations where complete population data are available. It does not involve the drawing of an inference from a sample to its population. For example, measures of central tendency—mean, mode, median, and measures of variability—standard deviation, average deviation and range—are descriptive statistical techniques.

Statistical procedures used for drawing of inferences about the properties of populations from sample data are generally referred to as *sampling or inferential techniques*. Inferences about population drawn from sample measures may involve some error or discrepancy, the magnitude of which can be estimated on the basis of the probability theory. Hence, mere description of the numerical properties of samples is within the realm of descriptive statistics while making inferences from

small to larger groups of subjects or events on the basis of probability theory in the province of inferential statistics.

1.5 Variables and Their Types

The term *variable* refers to a property or characteristic on which the members of a group or set differ from one another. These properties can be sex, age, grade, height, weight, intelligence, attitudes, socio-economic status and a host of other such factors. Opposed to the term *variable* is the term *constant* signifying the condition that the members of a group do not differ among themselves on this property. However, a particular property may be a variable in a specific situation and a constant in another situation. The property of sex, in a mixed group of boys and girls is a variable, while in a group of boys only, is a constant. The particular values of a variable are referred to as *variates* or *variate values*.

Variables may be *continuous* and *discrete* (discontinuons). A continuous variable may take an infinite number of values between any two points on the scale. Height, weight and chronological time are examples of continuous variables. A discrete variable can assume only a finite number of values between any two points on the scale. Size of family is a discrete variable. The theoretical nature of the variable and not its operations of measurement makes a variable continuous or discrete.

Variables may be classified into *Independent* and *Dependent* categories, if their functional relationship is of interest. The expression $Y=f(X)$ signifies that the given value Y is some unspecified function of another variable X . It shows that given a value of X and a knowledge of a functional relationship, Y can be predicted. In experiments, the *independent* variable is the stimulus variable in whose effect the experimenter is interested. The *dependent* or criterion variable appears, disappears or varies as the independent variable is introduced, withdrawn or given in different quantities.

1.6 Measurement Scales

Measurement can be defined as the assignment of numbers to objects and events according to logically acceptable rules.

The number system is highly logical and offers a multiplicity of possibilities of further logical manipulations. A measurement scale should possess the following attributes to allow for these logical manipulations.

Magnitude is the quantum or quantity in which the attribute exists in various instances of the phenomena. It allows us to tell whether one instance of the attribute is greater than, less than or equal to another instance of the attribute.

If X gets a score of 20 on an aggressiveness scale, and Y, a score of 25, we can say that Y is more aggressive than X.

Equal Intervals It denotes that the magnitude of the attribute represented by a unit of measurement on the scale is equal regardless of where on the scale the unit falls.

A difference in heights between 60 inches and 65 inches is equal to the difference in height, between 67 inches and 72 inches. However, when working with psychological phenomena, it may not be possible to interpret the equality of units at different points of the measurement scale. For example, a difference of IQ's between 170 and 190 may not be considered to be equal to the difference between IQ's of 100 and 120. Hence the IQ scale does not possess equal intervals. Another attribute of a measurement scale is an absolute zero point.

Absolute Zero Point is a value that indicates that a zero quantity of the attribute exists at that point or nothing at all of the attribute being measured exists.

For example, a zero height indicates "no height" at all and a zero weight "no weight" at all. However, in the case of intelligence and aggressiveness one may assign zero score to a person but it does not mean a point of an absolute absence of all intelligence or aggressiveness in the person.

Keeping in mind these three characteristics of measurements, the measurement scales can be divided into four different types.

1.6.1 Nominal or Classificatory Scale

It refers to the simple classification of objects or items into discrete groups which do not bear any magnitude relationships to one another. Generally numbering of houses, naming of streets, naming of persons and cars is done for convenience and

not based on any of the three qualities mentioned above – magnitude, equal intervals and an absolute zero point. Some people do not regard nominal scale as a scale at all.

1.6.2 Ordinal or Ranking Scale

It reflects only magnitude and does not possess the attributes of equal intervals or an absolute zero point.

For example, we may line up the students of a class according to height, and then instead of measuring them with a measuring tape, we merely rank them according to their height, the tallest receiving a rank of "1", the next tallest, a rank of "2", etc. We may ask the teacher to rank the students of his class on cleanliness, regularity, studiousness, etc. Clearly, the scale has magnitude but does not possess equal intervals. The distance between ranks "1" and "2" may not be equal to the distance between ranks "3" and "4", and so on.

Neither, it has the attribute of an absolute zero point; Even if we may assign a rank of "0" to any person, it does not mean that he possesses a zero level of that characteristic.

1.6.3 Interval Scale

The interval scale possesses two out of three important requirements of a good measurement scale i.e., magnitude and equal intervals but lacks the real or absolute zero point.

If we apply a meter scale to measure heights of the students in the instance mentioned above, we may find their heights to be 150 cm, 160 cm and so on. These measurements are on a scale with equal intervals. The distance of height between 150 cm and 160 cm is exactly equal to the distance of height between 170 cm and 180 cm. We can make some meaningful comparisons saying that Ram is 10 cm taller than Sham if their heights are 150 cm and 160 cm respectively. We can further say that object X is twice as long as object Y and so on. This type of scale allows the use of a large number of statistical analyses while the ordinal and nominal scales permit only limited analysis. The measurement of temperature in degrees Fahrenheit* is another example of interval scale.

*The Kelvin Scale of temperature does have an absolute zero point and thus can be considered as a ratio scale.

Table 1.1
Characteristics and Examples at a Glance of Measurement Scales

<i>Scale</i>	<i>Magnitude</i>	<i>Equal Intervals</i>	<i>Absolute zero point</i>	<i>Examples</i>
Nominal	Not Present	Not Present	Not Present	Class names; Room Numbers; Names of Persons; etc
Ordinal	Present	Not Present	Not Present	Ranks allotted to students on regularity, cleanliness; etc.
Interval	Present	Present	Not Present	Intelligence scores; Personality scores; etc.
Ratio	Present	Present	Present	All physical measurements, like height, weight, etc; Num- ber of students in various classes; Number of books possessed by students of a class; etc.

Although it allows for the comparison of intervals at different points of the scale, a zero degree F does not indicate an absolute absence of all heat. However, it allows for such statements that a particular day is twice as hot as another day when the temperatures are 30 F and 15°F respectively.

1.6.4 Ratio Scale

The scale of measurement which has all the three attributes – magnitude, equal intervals and an absolute zero point – is called a ratio scale. The interval scale discussed above does not have an absolute zero point at which a complete absence of a particular property can be taken. An absolute zero point, which is an additional characteristic of ratio scales means exactly nothing of the quantity being measured whether it is a physical or a psychological variable that is concerned. Ratio scales are almost non-existent in psychology and other social sciences, except in the area of psychological judgement. Ratio scaling allows for mathematical manipulations of multiplication, division and taking of square-roots, etc. and permit us to arrive at sensible results that can be verified. In statistical techniques, we create meaningful zero points such as at the mean of a distribution or at a difference of zero. Deviations from these statistically generated zero points can be treated as ratio scale measurements.

Exercises for Practice

- 1.1 What do you mean by the term 'Statistics'?
- 1.2 What are the uses of statistics to students of education?
- 1.3 Define the following and give examples :
 - (a) Statistics (b) Parameters
 - (c) Descriptive and inferential statistics
 - (d) Continuous and discrete variables.
- 1.4 For each of the following examples, give the highest level of measurement scale involved:
 - (a) Number of boys in a history class.

- (b) Number of kilograms of weight a boy can lift.
- (c) Temperature on a Centigrade scale.
- (d) Roll Numbers assigned to students for examination purposes.
- (e) Serial numbers assigned to lottery tickets.
- (f) Number of cars possessed by the residents of a city.
- (g) Ranking of students on personal hygiene.
- (h) Number of items answered correctly by a student in the examination.
- (i) Measurements on an attitude scale.

CHAPTER 2

FREQUENCY DISTRIBUTIONS AND THEIR GRAPHIC REPRESENTATION

2.1 Frequency Distributions

The data obtained from the conduct of experiments or surveys are frequently a collection of numbers or scores. A *frequency distribution shows a tallying of the number of times each score value (or interval of score values) occurs in a group of scores*. Classification and description of scores in the form of a frequency distribution have merits. The task of communication becomes easier and briefer and is better understood, and the important features of the data may become clearer. Frequency distributions make the manual calculations of statistics easier. However, the main *flaw* is that the individual scores lose their identity and take over the central value of their respective class interval. It may lead to some error. Frequency distributions may be unnecessary when computer or calculating machines are to be used.

Suppose a statistics test was given to 10 students and their scores noted. How can we arrange these scores into a frequency distribution? Table 2.1 shows the procedure for the same. In Part A, the scores obtained by the students have been given. For a better understanding of the scores, they have been listed in descending order (decreasing order) in Part B. It is evident that there are two values of 14, two of 13, three of 12, two of 11, and one of 10. We designate score value as "X" and the tally marks are made next to the X for each occurrence of the value and the tallies counted and written as frequency, f, against each score. The result is the frequency distribution

TABLE 2.1

The Development of a Frequency Distribution**A. Scores on a statistics test**

Given	12	12
	14	14
	13	10
	11	12
	13	11

**B. Scores presented in decreasing order
(Preparation of an Array)**

Step I	14	12
	14	12
	13	11
	13	11
	12	10

C. Putting tally marks and frequencies**Step II**

<i>Score</i>	<i>Tally</i>	<i>Frequency</i>
<i>X</i>		<i>f</i>
14	//	2
13	//	2
12	///	3
11	//	2
10	/	1

N=10

as presented in Part C of Table 2.1. This frequency distribution indicates the frequency for any given value of the variable X . Note that the sum of the frequencies is equal to N , the total number of values of X , i.e. 10. In the final form, the tally marks are deleted and only X and f values retained.

2.2. Relative Frequency Distribution

A distribution that indicates the percentage of the total number of cases which were observed at each score value (or interval of values) is called a relative frequency distribution. The frequency distribution given in Table 2.1 is reproduced in Table 2.2 to facilitate the calculation of relative frequency for each score and explain the procedure.

TABLE 2.2
Relative Frequency Distribution Based on Data
of Table 2.1(c)

Score X (1)	f (2)	Relative frequency (3)
14	2	.2
13	2	.2
12	3	.3
11	2	.2
10	1	.1
N=10		1.00

Frequency for each score (or class interval of scores) is divided by N and the proportions thus obtained are entered as relative frequency in column (3) of Table 2.2. The relative frequencies thus sum up to 1.00. Percentages instead of proportions may also be used. The advantage of a relative frequency distribution is that it expresses the pattern of scores in a manner which is not so dependent on the specific number of cases involved. Instead of saying that 3 students got a score of 12 each, one would say that 30 per cent did so. By breaking the dependency on N 's, two or more relative frequency distributions can be compared meaningfully which is not so convenient in the case of simple frequency distributions. Of course, it is always informative to know the total number of persons or N , in each case. The above example illustrating the concept of a

TABLE 2.3
One Hundred Hypothetical Scores

101	107	152	153	159	153	182	191	192	180
153	156	162	172	175	176	133	134	135	138
157	158	172	140	141	145	147	140	111	131
134	150	150	120	122	183	172	161	163	165
167	164	172	151	157	158	160	159	160	172
181	172	171	149	152	154	155	157	158	160
161	171	151	154	152	155	157	154	156	153
102	101	106	171	109	117	116	119	118	110
142	146	132	171	145	155	157	158	150	154
105	171	172	162	152	175	153	157	153	152

frequency distribution had a very small number of cases and a narrow range of scores. It was used to demonstrate through a simple case, the various aspects of this technique. However, frequency distribution is used to its greatest advantage where there are a large number of scores and a wide range of score values. In the formation of a frequency distribution, a decision regarding the size and number of class intervals is to be taken. It is to be based on some general principles and widely accepted conventions. The procedure of formation of a frequency distribution is illustrated on a set of 100 hypothetical scores given in Table 2.3.

The frequency distribution obtained from the above scores is given in Table 2.4 below :

TABLE 2.
Frequency Distribution of 100 Hypothetical Scores
with Class Interval (C.I.) size—10

(1) <i>Class Interval</i>	(2) <i>Tally</i>	(3) <i>f</i>
190-199	II	2
180-189	IIII	4
170-179	III III III	15
160-169	III III I	11
150-159	III III III III	38
140-149	III III III III	8
130-139	III III	6
120-129	III I	2
110-119	III II	7
100-109	III II	7
		N=100

2.3 Steps

(1) Determine Range

Range means the highest score minus the lowest score. In Table 2.3, the highest score is 192 and the lowest, 101. The range therefore is equal to $192-101=91$.

(2) *Decide about the Number and Size of Class Intervals*

A class interval (CI) is a band of scores. As a convention, the number of class intervals is generally kept between 10 and 20. However, depending upon the size of N , this number can vary beyond the two limits. In fact, the smaller number of class intervals leads to a coarser grouping and may introduce small errors into calculations of various statistics. The larger number of class intervals provides for a greater accuracy but may involve complications of calculations especially when N is not very large.

The sizes of class interval generally recommended are 1, 2, 3, 5, 10 and 20. These six sizes fit into all types of data generally used by the social scientists. In our case, we preferred to have a class interval size of 10 for our data. The number of class intervals thus would be: $\text{Range} \div \text{class interval size}$ i.e. $91 \div 10 = 10$ (raised to the next higher number). Had we decided to have a class interval size of 5, the number of class intervals would have been 21; and with a size of 3, thirty one.

(3) *Determine the Starting Point*

As a good rule, start the intervals with their lowest scores at multiples of the size of the interval. In our case, having CI size equal to 10, we started with 10, 20, 30, 40, etc.; for a size of 3, start with 3, 9, 12, 15, etc., and for a size of 5, start with 5, 10, 15, 20, 25, etc. Starting with an odd number when the CI size is also odd, gives an additional advantage that the mid-points of the class intervals will also be multiples of the CI size. In our case, we decided to start with 100, the lower limit of the lowest class interval. It is the first multiple of 10, below the smallest score in our data i.e. 101. The class intervals thus formed are shown in Table 2.4. Instead of writing all the scores like 100, 101, 102, 103, 104, 105, 106, 107, 108 and 109, we have written only the lowest and the highest limits i.e. 100-109. This is to save space and to avoid unnecessary lengthening of the table of frequency distribution. Moreover, the brief version, 100-109, means that all the ten scores between 100 and 109 are included. The class intervals as a convention, are written in a descending order, keeping the highest CI at the

top, and the lowest CI at the bottom. They should also form a continuous series and not the broken ones

(4) *Tally the Frequencies*

In Table 2.4, Col (2), the tally marks have been shown. Taking each score from Table 2.3, we locate it within the proper interval and write a tally mark " ", against the interval. Complete the tallying so as to cover all the scores. As a general convention based on convenience in counting, after four tallies have been placed before a CI, the fifth tally is marked from right to left across the first four so as to make clearly distinguishable groups of five tallies each. Tallies for each CI are then counted and noted as frequency for that CI (see Col. 3 of Table 2.4). The sum of frequencies should be equal to the total number of scores or N . In case any score has been omitted or duplicated in tallying, the total number of frequencies will not equal the total number of scores, or N . In case a tally has been placed in a wrong interval, there is no way of checking this error. However, in case of an error or doubt, the whole process of tallying should be gone over again.

The sum of frequencies is written as Σf in which Σ (capital Greek letter Sigma) stands for "the sum of"; and f , for frequencies. The total number of scores or individual is symbolized by the capital letter N which stands for number.

Summary of General Rules and Conventions

- 1 Select a class interval size that leads to the formation of 10 to 20 class intervals covering the whole range of scores.
- 2 As far as possible, select class intervals of size, 1, 2, 3, 4, 5, 10 or 20 points. These are likely to meet the requirements of most sets of data.
- 3 Start the class interval at a score which is multiple of the size of that interval.
- 4 Place the class interval containing the largest observations at the top, and that containing the smallest observations, at the bottom. All other class intervals should be arranged in descending order in between.

5. Make class intervals of uniform size and without any break in the series.

2.4 Exact Limits and Mid points of the Class Intervals

In the case of continuous variables, recording of observations or scores as discrete values like, 10, 11, 12, 13, etc is based on the assumption that the value recorded represents a value falling within certain limits. These limits are usually taken as one-half or .5 unit above and below the value reported. A score of 51 on a test of general intelligence would imply that if a more accurate form of measurement has been used, this value would fall within the limits 50.5 and 51.5. In more precise terms, these limits are 50.5 to 51.499, in which the latter value has a recurring decimal, but for convenience, we write the limits as 50.5 to 51.5. If the measuring process allows for a higher level of accuracy like measurement to the nearer thousandth of a second (As in psychophysical experiments), an observation of say, 28.8 seconds would fall within the limits of 28.45 and 28.85 seconds. These limits are often known as *exact limits* of the class interval. The terms, *real limits*, *class boundaries* or *end values* are also used for the purpose.

In the case of discrete variables, the class intervals themselves as given represent the exact limits and no subtraction and addition of one half unit is required to arrive at the exact limits. In Table 2.5, the exact limits of class intervals based on the data of Table 2.4 have been shown. For instance, the lowest class interval of 100-109 has the exact limits of 99.5, 109.5.

For this purpose, 99.5 the lowest exact limit of the score 100 has become the lower exact limit of the class interval. Similarly, 109.5, the upper exact limit of the score 109, has become the upper exact limit of the same class interval.

The *mid-point* of class intervals are required in the calculation of further statistics like, mean, median and standard deviation, etc for which the frequency distributions are generally constructed. Moreover, all the individual scores in a class interval lose their identity and take over the value of the mid-point of the class interval to which they belong.

It is not difficult to determine the mid-points of the class

TABLE 2.5

Exact Limits and Mid-points of Class Intervals

<i>Class Interval</i>	<i>Exact limits (Lower-Upper)</i>	<i>Mid-point of Intervals (X)</i>	<i>Frequency (f)</i>
190-199	189.5-199.5	194.5	2
180-189	179.5-189.5	184.5	4
170-179	169.5-179.5	174.5	15
160-169	159.5-169.5	164.5	11
150-159	149.5-159.5	154.5	38
140-149	139.5-149.5	144.5	8
130-139	129.5-139.5	134.5	6
120-129	119.5-129.5	124.5	2
110-119	109.5-119.5	114.5	7
100-109	99.5-109.5	104.5	7
			$\Sigma f = 100$

intervals. The mid-point of any class interval can be obtained by adding one-half of the range of the class interval to the lower exact limit of that interval. To illustrate the point, we may refer to the two lowest class intervals of the frequency distribution in Table 2.5. The CI has a size of 10 units.

<i>Class interval</i>	<i>Lower exact limit</i>		<i>One-half of range</i>	<i>Mid-point</i>
110-119	109.5	+	5	114.5
100-109	99.5	+	5	104.5

The important point is to take the lower exact limit and not the stated limit.

Generally, the midpoint of a class interval can also be obtained by using the formula.

$$\text{Mid-point} = (\text{Lower limit} + \text{Upper limit})/2 \quad \dots(2.1)$$

This applies to both the stated as well as the exact limits. In our case,

Mid-point of class interval $110-119 = (110 + 119)/2 = 114.5$
 and Mid-point of class interval $100-109 = (100 + 109)/2 = 104.5$

Using the exact limits also, we obtain the same values :

Mid-point of
 class interval $110-119 = (109.5 + 119.5)/2 = 114.5$

Mid-point of
 class interval $100-109 = (99.5 + 109.5)/2 = 104.5$

The concepts of size, exact limits and mid-points of a class interval are illustrated in Figure 2.1.

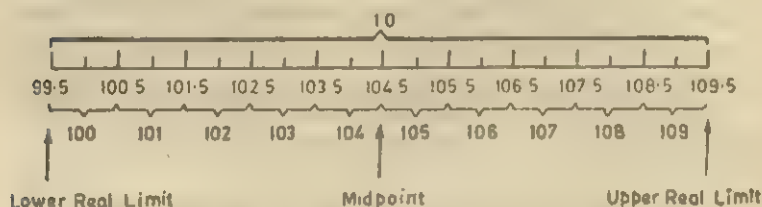


Fig. 2.1 Illustration of Size, Lower and Upper exact limits and Mid-point of the Interval 100-109.

It may seem a bit baffling that the size of the class interval 100-109 is 10 and not 9 as it might initially appear. But if all the scores contained in the class interval 100-109 are listed (100, 101, 102, 103, 104, 105, 106, 107, 108 and 109), these are clearly 10 and not 9 of them. Moreover, it may be kept in mind that both the scores of 100 and 109 are included in the interval. Hence, it is advisable that we do not calculate the size of the class interval simply by subtracting the stated lower limit from the stated upper limit of the interval. It would mean a size deficient by one unit. However, if the lower and upper exact limits are used for the purpose, the answer will be correct as in the case of the class interval 100-109:

$$\text{Size of class interval} = 109.5 - 99.5 = 10.$$

2.5 Assumptions regarding Values Within the Intervals

The grouping of data in class intervals leads to a loss of information about individual scores. Moreover, class intervals having identical lower and upper limits may be based on entirely different values as shown below. Let us take the class interval 5-9 for illustration:

	<i>Individual scores</i>	<i>Class interval</i>
1.	5,5,5,5,5	5-9
2.	5,6,7,8,9	5-9
3.	9,9,9,9,9	5-9

In case (1), all the scores are concentrated at the lower limit, while in case (3), at the upper limit of the class interval. In case (2), the scores are evenly spread over the entire range of the class interval. Many other combinations of these scores are possible which can lead to the formation of this class interval.

Hence, in the calculation of various statistics, two important assumptions about values within the intervals are made.

1. The first assumption which is generally used in the calculation of statistics like, median, quartiles, percentiles, etc. is that *the observations or scores are uniformly distributed over the entire range of the interval.*
2. In the calculation of means, standard deviations and in drawing frequency polygons, another assumption is made that *all the values or scores in the class interval are the same and equal to the value corresponding to the mid-point of the interval.*

2.6 Graphic Representation of Data

A graph is the geometrical image of a frequency distribution. It is a mathematical picture. Frequency distributions are converted into visual models to facilitate understanding. It is easier, more convenient and quicker to draw inferences from graphs than from frequency distributions. Comparison

2. *Abscissa* is the X-axis or XOX line along which the X values are located. It is the horizontal line in Figure 2.2.
3. *Ordinate* is the vertical line or YOY' line or Y-axis along which Y values are located.
4. *Origin*, the point of origin or the zero point is the point of intersection of XOX' and YOY'.
5. *Quadrants*. The whole area of the plane is divided into four quadrants showing the signs of the values of X and Y as below:

Quadrant I : X positive and Y positive

Quadrant II : X negative and Y positive

Quadrant III : X negative and Y negative

Quadrant IV : X positive and Y negative.

6. *Signs of values:*

On ordinate : signs above O, positive; and
signs below O, negative.

On Abscissa : signs to the right of O, positive; and
to the left of O, negative.

7. By convention, the scores are laid off on horizontal or X axis and the frequencies on the vertical or Y axis.
8. It is customary to assign a self-explanatory title to every graph.
9. The distance along either axis selected to serve as a unit should be such that ratio of height to length is roughly 3:5. It adds to the aesthetic appearance of the graph.

Now a description of the following will be presented:

1. Histograms
2. Frequency Polygons
3. Cumulative Frequency Curves

2.6.1 Histograms

A histogram is a set of vertical bars with equal base but different heights. Therefore, it is also known as *bar-graph*. It

is known as *frequency histogram* also. The mechanics of its construction will be explained with reference to the data of Table 2.6, plotted as a histogram in Figure 2.3.

TABLE 2.6

**Cumulative Frequency and Cumulative Percentage
Frequency in a Frequency Distribution**

(1) <i>Class interval</i>	(2) <i>Exact limits</i>	(3) <i>Frequency</i>	(4) <i>Cumulative frequency</i>	(5) <i>Cumulative percentage frequency</i>
45-49	44.5-49.5	2	50	100.00
40-44	39.5-44.5	3	48	96.00
35-39	34.5-39.5	6	45	90.00
30-34	29.5-34.5	9	39	78.00
25-29	24.5-29.5	13	30	60.00
20-24	19.5-24.5	8	17	34.00
15-19	14.5-19.5	6	9	18.00
10-14	9.5-14.5	2	3	6.00
5- 9	4.5- 9.5	1	1	2.00

N=50

In Figure 2.3, the baseline is labelled with the score intervals rather than with the exact limits. Thus, the first interval in the histogram actually begins at 4.5, the exact lower limit of the interval, and ends at 9.5, the exact upper limit of the interval. The one score or frequency in the interval 5-9 is represented by a rectangle, the base of which is the length of the interval and the height of which is one unit up on the Y axis. The two scores or frequencies on the next interval, 10-14, are represented by a rectangle with a length of one interval and height of 3Y units. The highest rectangle is on interval 25-29, which has a frequency of 13. The numbers written at the top of each rectangle will in the initial stage facilitate reading of the frequencies.

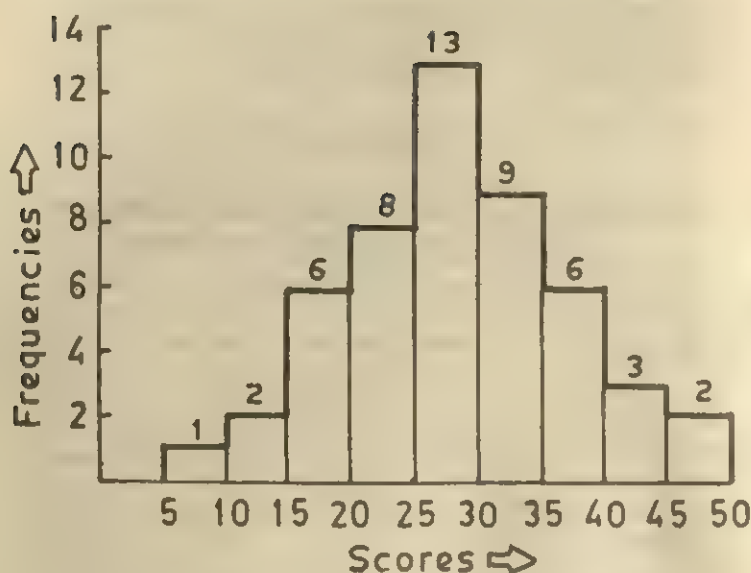


Fig. 2.3. Histogram of the 50 Scores given in Table 2.6.

However, these are not always necessary. In selecting scales for the X axis and Y axis, the consideration of an approximate ratio of 3:5 between the height and length should be kept in view.

The histogram is composed of rectangles with different heights. It is not necessary to project the sides of the rectangles down to the base as is done in Figure 2.4. Still it will bring out the important fact of the rise and fall of the frequencies from interval to interval.

2.6.2 Frequency Polygons

A polygon is defined as a many-sided figure. When a many-sided figure is drawn on the basis of frequencies given in a frequency distribution, the figure is called a frequency polygon. Construction of all graphic figures requires the selection of a good graph paper with cross-sections. For polygons, a graph paper ruled into heavy lines 1 inch apart each way, and subdivided into tenths of an inch more lightly drawn will be more convenient.

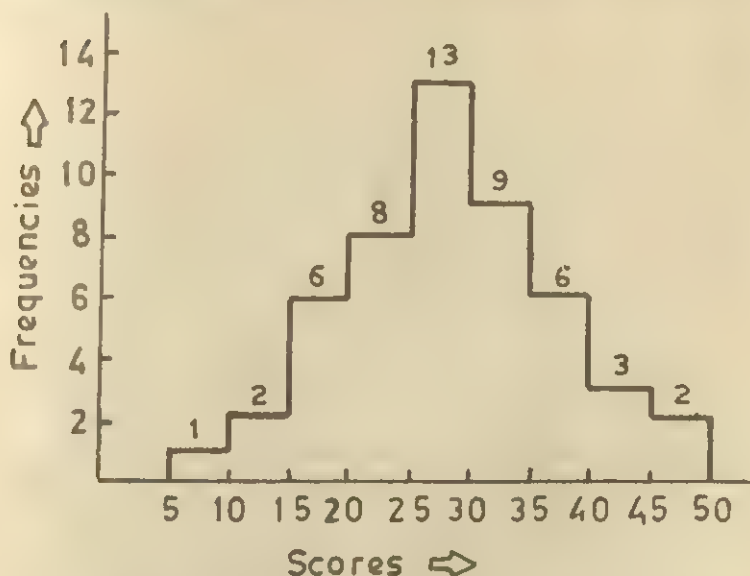


Fig. 2.4. A Histogram with Sides of Rectangles not Projected to the Baseline (Data in Table 2.6)

Since a polygon is a complete figure, its ends should touch the baseline. For this purpose, at each end of the distribution, assume one additional class interval with zero frequency. In Table 2.6, there are in all 9 class intervals and with the assumption two additional class intervals, this number will go up to eleven. By allowing $1/2''$ to each class interval, the distribution will spread over an extent of $5\frac{1}{2}''$ in. which is sufficiently large for easy readability. Since there are five scores in each class interval, the total units would be $5 \times \text{No. of class intervals}$ i.e., $5 \times 11 = 55$. Hence $1/5''$ of space (two subdivisions) will be allowed to each unit. On the base line, every fifth line will be labelled with a multiple of five such as 0, 5, 10, . . . , 50, 55.

The height of the figure should be roughly $3/5$ to $3/4$ or 60%-75% of the total width. Our total width is $5.5''$. Hence, a height of $5.5'' \times 3/5$ to $5.5'' \times 3/4$ (or $3.3''$ to $4.2''$) will be appropriate. We may choose $4.2''$ as it will be divisible by the maximum frequency of 13 in our data. Hence, four small

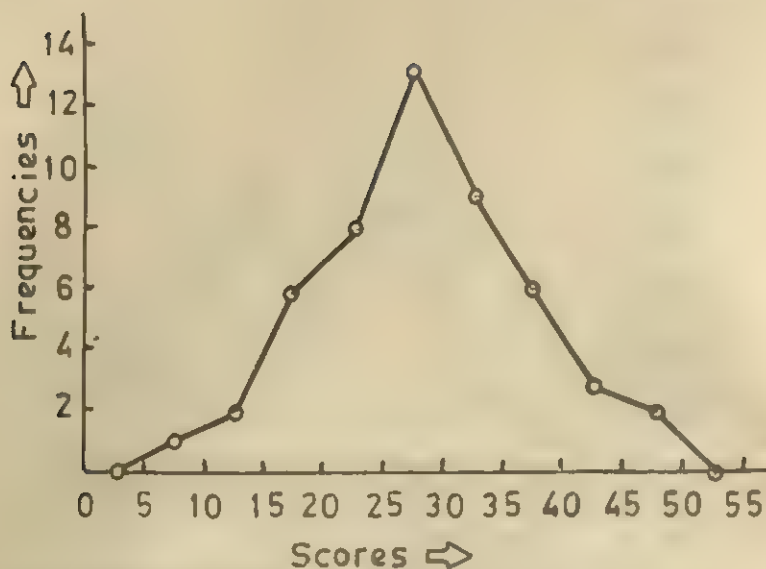


Fig. 2 5. A Frequency Polygon for the Data in Table 2 6.

squares ($2/5$ in.) on the Y-axis will represent one unit of frequency. This will satisfy the general convention about the ratio of width to the height of the polygon.

The next step is to locate the mid-points of the class intervals. It can be done by averaging either the exact or the stated limits of each class interval. In our case for class interval, 5-9, the mid-point is $(5+9)\frac{1}{2}$ or 7; and for 10-14, it is $(10+14)\frac{1}{2}$ or 12.

Now we have to plot the dots for the frequency polygon. For the first class interval (additional one), 0-4, the frequency is zero. Hence, the dot is placed at the mid-point of the class interval on the baseline. For the next class interval of 5-9, the dot is placed exactly above the score 7 and at a perpendicular distance from the relevant frequency of 1. Dots for the other class intervals are to be plotted in the same manner keeping the relevant frequency in view. The dot for the last (additional) class interval will be on its mid-point on the baseline.

Now join the dots with straight lines. The curve so drawn is the frequency polygon as shown in Figure 2.5.

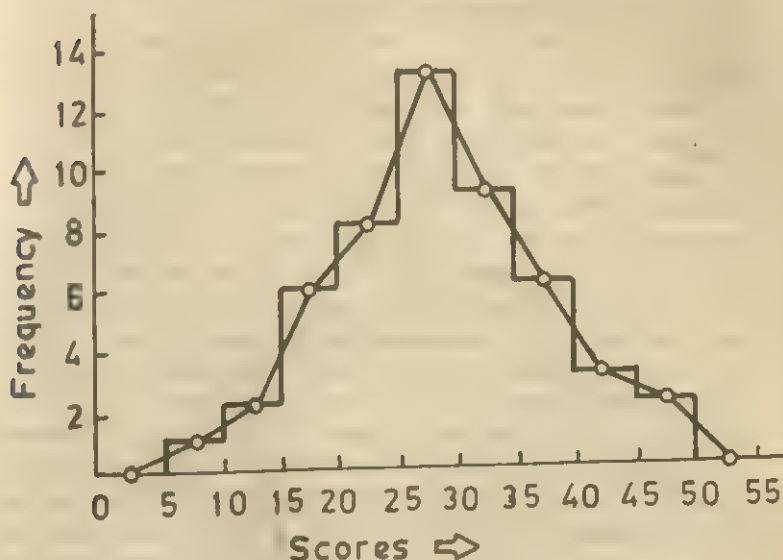


Fig. 2.6. A Frequency Polygon Constructed from a Histogram given in Figure 2.4.

A frequency polygon can also be constructed by joining the mid-points of horizontal lines of a histogram. Figure 2.6 illustrates the procedure of the same.

A frequency polygon is generally preferred to a histogram because of several reasons: (i) The former gives much better conception of the contour of the distribution. (ii) Another important merit is that it gives a more accurate impression that the cases are more frequent near the central tendency, mode. (iii) We can also plot two or more polygons overlapping on the same baseline for the purpose of comparison. If N in each case is unequal, the frequencies of both the distributions should be converted into percentages by multiplying each frequency by $100/N$. Here, N stands for the total number of cases in the relevant distribution. If N 's are equal, conversion into percentages is not required and the frequencies can be plotted straightaway.

2.6.3 Smoothed Frequency Polygon

When the sample is small and frequency distribution somewhat irregular, the frequency polygon tends to be jagged in its shape. Hence, with a view to iron out chance irregularities and obtain a better picture of how the figure would look like if the data were more numerous, the frequency polygon may be smoothed. The process involves taking of "moving" or "running" averages to determine the smoothed frequencies which are later plotted to form the smoothed polygon. In Table 2.7, Col. (4), the process of taking the "running averages" and the smoothed frequencies so obtained have been shown. To find out "smoothed" or "adjusted" frequencies, we add the f on the given interval and the f 's on the two adjacent intervals (the interval just below and the interval just above) and divide the sum by 3. To find the smoothed f 's for the two extreme intervals, namely 5-9, and 45-49 it is presumed that there are zero frequencies below the interval 5-9, and also above the interval 45-49. In Figure 2.7, a frequency polygon based on the original frequencies and another based on smoothed frequencies have been given.

TABLE 2.7
Smoothed Frequencies

(1) <i>Class Interval</i>	(2) <i>Mid-points</i>	(3) <i>f</i>	(4) <i>Smoothed f</i>
45—49	47	0	$(3+0+0)/3=1.00$
40—44	42	3	$(6+3+0)/3=3.00$
35—39	37	6	$(12+6+3)/3=7.00$
30—34	32	12	$(13+12+6)/3=10.33$
25—29	27	13	$(8+13+12)/3=11.00$
20—24	22	8	$(6+8+13)/3=9.00$
15—19	17	6	$(2+6+8)/3=5.33$
10—14	12	2	$(0+2+6)/3=2.67$
5—9	7	0	$(0+0+2)/3=0.67$
N=50			50.00

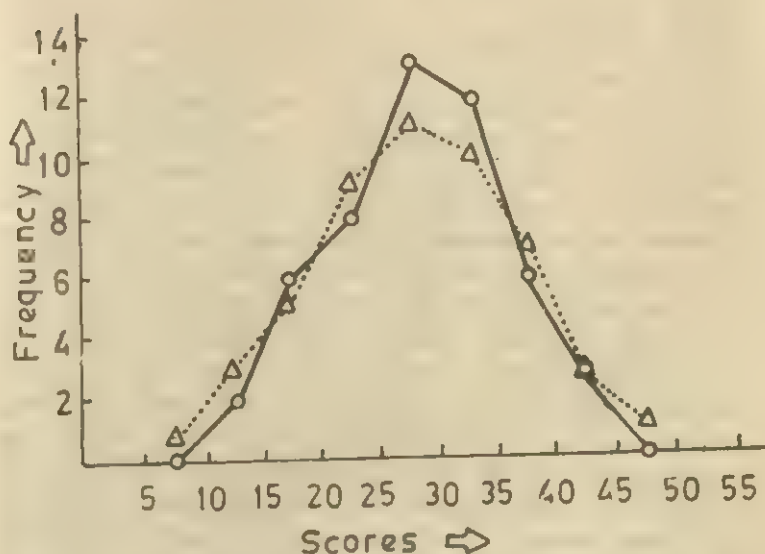


Fig. 2.7. Original and Smoothed Frequency Polygons based on data in Table 2.7.

Smoothing can be done twice to improve the outline of the frequency polygon and make it more flowing. However, so much adjustment of frequencies is seldom warranted. The original polygon must also be presented along with the smoothed one so that the extent of adjustment can be *gauged* by the reader. If N is large, smoothing may not largely improve the shape of the polygon. Moreover, smoothing is desirable with continuous variables; or distributions based on small samples, and the total area in both the polygons should be equal.

2.6.4 Cumulative Frequency Curve

In Table 2.6, Col. (3), frequencies belonging to different class intervals have been shown. In this section, we are interested in the frequencies falling below various score points on the measuring scale. *The cumulative frequency corresponding to any class interval is the number of cases within that interval plus all the cases in intervals lower to it on the scale.* In Table 2.6, cumulative frequencies have been shown in Col. (4). The method of calculating cumulative frequencies is very simple and

requires successive additions of ordinary or non-cumulative frequencies. The cumulation starts from the bottom. For example, in our case, there is a frequency of 1 in the lowest interval of 5-9; and there are no frequencies below it. Hence, the cumulative frequency for this class interval will be $1+0=1$ (See Col. 4 of Table 2.6). For the next higher interval, 10-14, the cumulative frequency will be 1 plus 2=3. For the third interval, it would be 6 (the frequency of the interval) plus 3 (the cumulative frequencies immediately below it), equal to 9. The process is to be continued till we reach the top interval for which the cumulative frequency will always be equal to N . In case of a discrepancy between the two, some error is sure to have crept in. It is very important to understand the meaning and interpretation of a cumulative frequency. In our example, the cumulative frequency of 9 in the third class interval of 15-19 means that 9 cases fall below the exact upper limit, i.e., 19.5 of the class interval. The top cumulative frequency of 50 shows that all the 50 cases fall below the exact upper limit of the top interval, i.e. 49.5.

The drawing of a *cumulative frequency curve* differs from that of a frequency polygon in two respects:

- (i) Instead of plotting points above the mid-points of the class intervals, we plot them above the exact upper limits of the class intervals.
- (ii) Instead of using ordinary frequencies for plotting the points, we use cumulative frequencies.

In plotting the cumulative frequency distribution of Table 2.6, we would plot the cumulative frequency of 1 above the exact upper limit 9.5, the lowest class interval, 5-9. The cumulative frequency of 3 would be plotted over the exact upper limit, 14.5 of the next higher class interval, 10-14. For this purpose, the exact upper limits, instead of the mid-points, are to be marked on the baseline. A cumulative frequency curve based on the data of Table 2.6 has been plotted in Figure 2.8.

It may be noted that the general trend of the cumulative frequency curve is progressively rising; there are no inversions or setbacks. The upward rise is not a straight line. When the

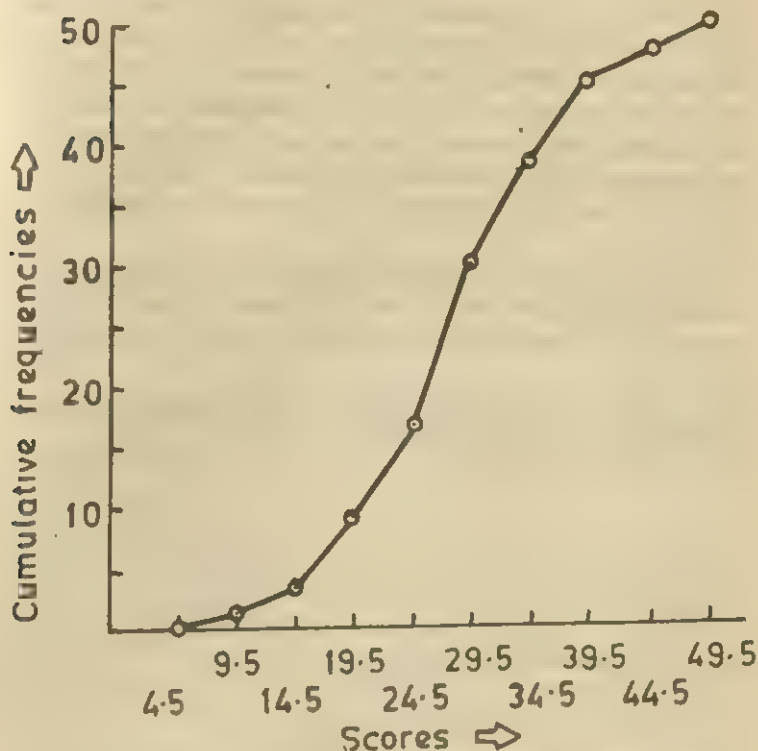


Fig. 2.8. A Cumulative Frequency Curve based on the Data in Table 2.6.

distribution of frequencies is symmetrical, the cumulative distribution curve is usually S-shaped. The zero frequency does not lead to any inversion in the curve but may show up as a plateau where the height of the curve remains constant.

2.6.5 Cumulative Percentage Curve or Ogive

There are occasions when cumulative frequencies become more meaningful and convenient, when converted into cumulative percentages. This process makes a comparison of two or more distributions possible, when N differs. This leads to a standardization of N at 100. The determination of percentiles and percentile ranks becomes possible and need of calculation is eliminated (Also see Chapter 5).

In Table 2.6, Col. (5) cumulative percentages corresponding to various cumulative frequencies have been given. The process of conversion is simple. Cumulative percentage for any cumulative frequency can be obtained by multiplying the *latter* by $100/N$. For example, in our example, $N = 50$. For the cumulative frequency of 1, the cumulative percentage would be $1 \times 100/50 = 2$. For the third interval, the cumulative percentage is $9 \times 100/50 = 18$.

The procedure of drawing an ogive is similar to that of a cumulative frequency curve except in one respect. In the former, we use cumulative percentages instead of cumulative frequencies.

A cumulative percentage curve or ogive based on the data of Table 2.6 is shown in Figure 2.9.

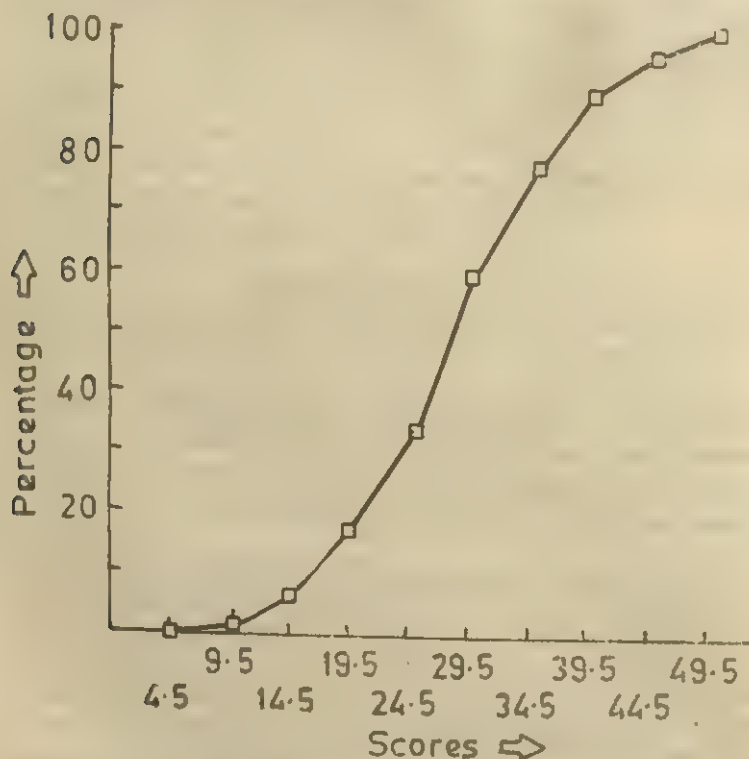


Fig. 2.9. A Cumulative Percentage Curve or Ogive based on the Frequency Distribution in Table 2.6.

Exercises for Practice

- 2.1 What is a frequency distribution? What are its uses in statistical analysis and presentation of data?
- 2.2 From the 50 scores given below, construct a frequency distribution keeping in view the various conventions and principles:

22, 12, 18, 23, 10, 9, 8, 50, 18, 17, 16, 30, 32, 33, 28,
21, 24, 30, 40, 42, 44, 46, 19, 10, 11, 15, 16, 31, 37,
36, 24, 25, 26, 28, 23, 21, 22, 8, 7, 6, 5, 7, 6, 9,
11, 30, 26, 27, 25, 26.

(Hints: Choose a class interval of size 5 and start with the first class interval of 5-9 and check your results with Table 2.6).

- 2.3 From the frequency distribution obtained in Question No. 2.2 above, prepare a smoothed frequency polygon.
- 2.4 Define the following as precisely as you can:
(i) Histogram, (ii) Frequency Polygon, (iii) Ogive.
- 2.5 Compare histogram and frequency polygon regarding their usefulness in graphical representation of data.
- 2.6 What are the uses of an Ogive?
- 2.7 Draw a histogram from the data given below and superimpose a frequency polygon upon it:

<i>Scores</i>	<i>f</i>	<i>Scores</i>	<i>f</i>
190—199	2	140—149	8
180—189	4	130—139	6
170—179	15	120—129	2
160—169	11	110—119	7
150—159	38	100—109	7

- 2.8 Draw a cumulative frequency curve and an ogive from the data given in Q. No. 2.7 above.
- 2.9 Draw a smoothed frequency polygon from the data in 2.7 above.

CHAPTER 3

MEASURES OF CENTRAL TENDENCY

Imagine an obtained distribution of numerical scores. If you were asked to state one value that would best "capture" and communicate the distribution as a whole, which value should you choose? One way to answer this question is to find that score value which is a good "bet" about any randomly selected case for the distribution. Such a score may not be exactly correct for any given case but it should be a fairly good guess about the obtained score for that case. However, there are three different ways to specify what we mean by a "good bet" about any case:

- (i) The arithmetic average of the distribution.
- (ii) The point exactly midway between the top and bottom halves of the distribution, and
- (iii) The most frequently occurring score or the mid point of the most frequent measurement class.

The first of these ways of defining the central tendency leads to the familiar measure known as the average or the *mean*; the second leads to the *median* of the distribution; and the third is known as the *mode*. All these three, as a class, are known as measures of central tendency. Though, the "average" is the popular term for the arithmetic mean, yet in statistical work "average" is the general term for any measure of central tendency.

3.1 The Mean (M)

The most used and familiar index of central tendency for a

set of raw data or a distribution is the mean. *The mean is a simple arithmetic average.* It is a common place knowledge that to take the average of a set of raw scores, *we simply add all the scores up and divide by the total number of scores, N.* Consider the following scores or measurements: 8, 14, 23, 10, 12, 5. The sum of these scores is 72. The arithmetic mean is, therefore, 72 divided by 6, 12 . In general, if N measurements are represented by the symbols, $X_1, X_2, X_3, \dots, X_N$, the arithmetic mean in algebraic language is

$$M = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N} \quad (3.1)$$

The Greek letter sigma $\sum_{i=1}^N$ describes the operation of summing the N measurements. The summation extends from $i=1$ to $i=N$. Generally the arithmetic mean is written simply as

$$M = \frac{\sum X}{N} \quad (3.2)$$

The limits of the summation are omitted. The summation is understood to extend over all available values of X. Sometimes \bar{X} (bar upon X) is used to denote the mean of X series; Y to denote the mean of Y series. In this text, M will be used as a generalized term for mean. However, for a distinction between the two different series of scores, M_x, M_y , etc. will be used. The student should note another fact about the mean:

$$M = \frac{\sum X}{N}$$

By cross multiplication, we obtain.

$$\sum X = NM \quad (3.3)$$

Thus the sum of a variable X is N times the mean of X. It is a useful concept and is used in a variety of situations.

Calculation of Mean from Frequency Distributions

When scores or measurements have been arranged in the form of a frequency distribution showing class intervals and frequencies, the following methods are used:

3.1.1 Calculation of Mean by Long Method

TABLE 3.1

Calculation of Mean from Frequency Distribution
with Class Interval of size 1.

Class Interval X	Frequency f	fX	
16	2	32	Computation Formula $M = \frac{\Sigma fX}{N}$ $\Sigma fX = 266$ $N = 20$ $M = \frac{266}{20}$ $= 13.3$
15	3	45	
14	4	56	
13	5	65	
12	3	36	
11	2	22	
10	1	10	
	20	266	

Symbols

Σ = Sum of

f = frequency

X = score

N = Total No. of scores

In the above frequency distribution, the class interval is of size 1. Each X is then multiplied by the relevant frequency and a product of the two, fX , obtained. If each X is denoted by X_1, X_2, \dots and frequency by f_1, f_2 etc. upto X_K and f_K respectively, the formula for the mean would be:

$$M = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \dots + f_KX_K}{N} = \frac{\sum_{i=1}^K f_iX_i}{N}$$

For simplicity, the limits of the summation are omitted and the formula becomes:

$$M = \frac{\sum fX}{N} \quad (3.4)$$

The calculation of mean from a frequency distribution with class interval size equal to more than 1 is also done in a similar way. It is shown below in Table 3.2.

TABLE 3.2

**Calculation of Mean from Frequency Distribution
with Class Interval size of two or more.**

<i>Class Interval</i>	<i>Mid-point</i>	<i>Frequency</i>	<i>Frequency × Mid point</i>
	(X)	(f)	(fX)
45—49	47	2	94
40—44	42	3	126
35—39	37	2	74
30—34	32	6	192
25—29	27	8	216
20—24	22	8	176
15—19	17	7	119
10—14	12	5	60
5—9	7	9	63
		N=50	$\sum fX=1120$

Formula

$$\text{Mean} = \frac{\sum fX}{N}$$

Substituting the values in the formula:

$$\text{Mean} = \frac{1120}{50} = 22.40$$

The steps in the calculation of mean by the long method are as follows:

1. Calculate the mid points of each class interval and call them X (Col. 2),
2. Multiply f and X (Col. 2 x Col. 3) to obtain fX (Col. 4),
3. Add fX values in Col. 4 to obtain ΣfX ,
4. Divide this sum, ΣfX , by N to obtain the mean.

3.1.2 Calculation of Mean by the Short Method or Assumed Mean Method

The long method of calculating mean as shown above is an accurate and straightforward method. However, it very often involves the handling of large numbers and requires tedious calculations. Hence, to overcome these difficulties, the "Assumed Mean" method or simply the Short Method has been devised for the computation of mean from the frequency distribution. The same is illustrated below:

TABLE 3.3

Calculation of mean from a Frequency Distribution by Short Method or Assumed Mean Method.

(1) <i>Class Interval</i>	(2) <i>Mid-point</i>	(3) <i>frequency</i>	(4) <i>Deviation of (X) from AM in units of C.I.</i>	
	(X)	(f)	(x')	(fx')
45—49	47	2	+4	+8
40—44	42	3	+3	+9
35—39	37	2	+2	+4
30—34	32	6	+1	+6
25—29	27	8	0	0
20—24	22	8	-1	-8
15—19	17	7	-2	-14
10—14	12	5	-3	-15
5—9	7	9	-4	-36
N=50			+27	
			-73	
			$\Sigma fx' = -46$	

Assumed Mean, $AM=27$; $\Sigma fx' = -46$

$$\text{Correction, } C = \frac{\Sigma fx'}{N} = \frac{-46}{50} = -.92$$

Size of class interval, $i=5$

Formula

$$\begin{aligned} \text{Mean} &= AM + Ci & (3.5) \\ &= 27 + (-.92 \times 5) \text{ (substituting the values)} \\ &= 27 - 4.6 = 22.40 \end{aligned}$$

In Short Method, we “guess” or assume a mean, and later apply a correction to the Assumed Mean (AM) in order to obtain the actual mean.

The steps involved in the calculation of Mean by the Short Method are as below:

1. Tabulate the scores into a frequency distribution and find out the midpoint of the class intervals (Cols. 1-3)
2. Take the midpoint of an interval somewhere near the centre of the frequency distribution and, if possible, the interval should contain the largest frequency. This is for convenience of computation as it would involve working with smaller values (If any other class interval is assumed, the value of mean will remain the same). In our example, the class interval, 25-29, is considered and its midpoint, 27, is taken as “assumed mean”.
3. The x' values as in Col. 4 are the *deviations from the assumed mean in units of class interval*. The midpoint of class interval, 25-29, is 27 and deviates 0 units from the AM and hence a zero is placed in the Col. x' against this interval. As we go up, we find the class interval, 30-34, deviates +1 unit, from the assumed mean. ($x' = \frac{\text{Midpoint} - AM}{\text{Size of CI}} = \frac{32-27}{5} = +1$). This value is placed opposite to this class interval in the x' column. The process is repeated to obtain values of +2, +3 and +4 for other class intervals above the class interval of 30-34. Let us come back to the class interval 25-29, which was assigned $x'=0$. The value of

x' for the class intervals below it are calculated in the manner as shown below:

CI	x'
20—24	$\frac{22-27}{5} = -1$
15—19	$\frac{17-27}{5} = -2$ and so on

The other x' 's thus are -3 , -4 and -5 .

However, the student will be able to see the simple way in which x' values can be assigned almost mechanically. Starting with $x'=0$ for the class interval having the AM, go up assigning x' values of $+1$, $+2$, $+3$ etc. till you reach the uppermost class interval. Once again starting from $x'=0$, go down assigning x' values of -1 , -2 , -3 , etc. till you reach the lowest class interval. This is possible because all class intervals are of uniform size.

4. fx' in Column 5 is the product of f and x' (Col. 3 \times Col. 4). While multiplying f and x' values the algebraic sign is to be kept intact and noted along with the values in Column 5. It may be noted that all fx' values in intervals above the AM are positive; and all fx' values below the AM are negative.
5. Obtain $\Sigma fx'$, the sum of fx' values by summing up algebraically the fx' values in Col. 5. For convenience, sum up separately the negative values and the positive values and obtain the absolute difference of the two sums. Give this difference, the sign of the larger sum. In our example, sum of x' values with plus sign is 27; that of minus sign is -73 ; the difference of the two sums is -46 with sign of the larger sum kept intact.
6. The value of correction, C , is obtained by dividing $\Sigma fx'$ by N . In our example, -46 is divided by 50 to obtain $C = -.92$ (the sign of the value is very important). Multiply C by i to obtain $Ci = -.92 \times 5 = -4.6$. Substituting these values in the formula, $M = AM + Ci$, we obtain $M = 27 + (-4.6) = 22.40$.

3.1.3 Some Properties of Mean

The mean possesses several properties which make it very useful. Some of them are described below:

1. The Mean as a "Center of Gravity" of a Distribution

The mean of a distribution parallels the physical idea of a center of gravity, or balance point, of ideal objects arranged in a straight line. For example, imagine an ideal board having zero weight. Along this board are arranged stacks of objects at various positions. The objects have uniform weight and differ from each other only in their positions on the board. The board is marked off in equal units of some kind, and each object is assigned a number according to its position as shown below in Figure 3.1.

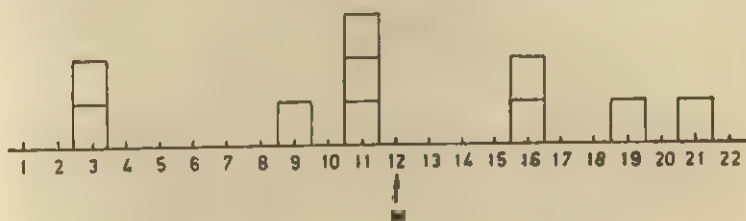


Fig. 3.1. Mean as Centre of Gravity of a Frequency Distribution.

Now given this idealized situation, at what point would a fulcrum placed under the board create a state of balance? That is, what is the point at which the "push" of objects on one side of the board is exactly equal to the push exerted by the objects on the other side? This is found from the mean of the positions of the various objects.

$$M = \frac{3+3+9+11+11+11+16+16+19+21}{10} = 12.$$

Here, the board would exactly balance if a fulcrum were placed at a position marked 12. It may be noted that this centre of gravity has been found in exactly the same way as for the mean of a distribution. The position of an object of uniform weight (midpoint of the class interval) was, in effect, multiplied by the number of objects at that position (the frequency). These

values were summed and divided by the number of objects (the total frequency or N).

2. *Deviations in one Direction of the Mean exactly equal the Deviations in the other Direction:*

This characteristic emerges from property No. 1 of the mean as mentioned above. Since mean is the "centre of gravity" or the "balance point" the sum of deviations in one direction of the mean exactly equals the sum of the deviations in the other direction. This leads us to a further conclusion that the *sum of deviations about the mean in any distribution is always zero*. This is illustrated below in Table 3.4.

TABLE 3.4
Deviations from the Mean

Score (Y)	Deviations (x)	Squares of Deviations (x^2)
6	+2	4
5	+1	1
4	0	0
3	-1	1
2	-2	4
<hr/> Σ $X=20$ $M=4$	<hr/> Σ $x=0$	<hr/> Σ $x^2=10$

It can also be shown that the sum of deviations taken from any other arbitrary value will not be zero. Σ x for arbitrary means of 3 and 6 are +5 and -10 respectively.

3. *The Principle of Least Squares*

Another property that emerges from the first two mentioned above is that the *sum of the squared deviations of all the scores about the mean is less than the sum of the squared deviations about any other value*. This is called the principle of least

squares. For example, in the above illustration (Table 3.4), the sum of the squared deviations about the mean equals 10. The M was 4. If 3 and 6 are taken as arbitrary values of the mean, the value of Σx^2 becomes, 15 and 30 respectively. Thus the sum (10) taken about the mean is less than any other of these examples, and it can be shown that it always will be less than about any other value. Hence, the essential property of mean is that *it is closer (in terms of squared deviations) to the individual scores over the entire group than is any other single value*. This is a highly useful concept that enters into several other statistical methods like regression and prediction. If we were told to guess the score of some case picked at random from a distribution, we may guess the mean for that and every other case so picked. It may not be true that the mean is exactly the same as any obtained. The overall signed error, on the average, will be zero and the sum of the squared errors will be the least.

4. The mean has the property that for most distributions, it is a more accurate, or more efficient estimate of the population mean than any other measure of central tendency, such as the median and mode, one of the population values they purport to estimate. It is subject to less error.

TABLE 3.5

Effect of a Constant on Mean

	Original score	Adding 2 to each score	Subtracting 2 from each score	Multiplying each score by 2	Dividing each score by 2
	4	6	2	8	1
	5	7	3	10	2.5
	6	8	4	12	3
	7	9	5	14	3.5
	8	10	6	16	4
Σ	30	40	20	60	15.0
M	6	8	4	12	3.0
Effect		(6+2)	(6-2)	(6×2)	(6÷2)

5. If a constant is added to each score of a distribution, the value of the mean will increase by the value of that constant. The subtraction of a constant from each score of a distribution will lead to a decrease in the mean equal to that constant. The multiplication and division will also lead to a result equal to the product of mean and the constant; and the quotient obtained by dividing the mean by that constant, respectively. An algebraic proof of this will be attempted later. However, a numerical example is given on p. 43 to demonstrate the above.

6. *The Combined Mean, M_{comb}*

A combined mean for two or more samples can be calculated if the M 's and N 's of those groups are available. This would avoid the necessity of combining the raw scores of all the samples and calculating the average in the normal way. For example, see Table 3.6.

TABLE 3.6
Calculation of Combined Mean

Group	K	M	Symbols
I	40	50	M_{comb} = Weighted arithmetic mean obtained from combining n groups. $N_1, N_2, \text{ etc.}$: No. of cases in groups I, II, etc. $M_1, M_2, \text{ etc.}$: Means of Groups I, II, etc.
II	30	45	
III	20	35	

Formula

$$M_{comb} = \frac{N_1 M_1 + N_2 M_2 + \dots + N_n M_n}{N_1 + N_2 + \dots + N_n} \quad (3.6)$$

substituting the numerical values

$$\frac{(40 \times 50) + (30 \times 45) + (20 \times 35)}{40 + 30 + 20} = \frac{2000 + 1350 + 700}{100} = \frac{4050}{100} = 40.5$$

When only two groups are involved, the formula becomes

$$M_{comb} = \frac{N_1 M_1 + N_2 M_2}{N_1 + N_2} \quad (3.7)$$

When N is equal, the formula reduces to

$$M_{\text{comb}} = \frac{M_1 + M_2 + \dots + M_n}{n} \quad (3.7a)$$

(where n is the number of groups)

For example, for the three groups having means of 10, 15 and 35 respectively and equal N 's the $M_{\text{comb}} = (10 + 15 + 35) / 3 = 20$

3.2 The Median (Md)

The median, symbolized by Md, is the point that divides the distribution into two parts such that an exactly equal number of scores fall above and below the point. It means that 50 per cent of the scores will be above the median and the remaining 50 per cent below it.

Computational examples. The computation of the median varies under different circumstances. The same is given below.

3.2.1 Ungrouped Data

(i) *When there is an odd number of scores in the distribution*

From distributions which have an odd number of scores but no duplication of scores near the median, the median is the middle score. The series is to be arranged in an ascending or descending order. For example Consider the distribution 6, 4, 8, 7, 10. Arranging the scores in ascending order 4, 6, 7, 8, 10, we find that the middle value 7, is the median. It divides the distribution into two equal halves.

(ii) *When there is an even number of scores in the distribution*

When there is an even number of scores and there is no duplication of scores near the median, the average of the middle two scores is taken as the median. The scores must be arranged in ascending or descending order.

For example. Consider the distribution 6, 4, 8, 7, 10, 5.

Arranging the scores in ascending order 4, 5, 6, 7, 8, 10

$$\text{Median} = \frac{\text{Sum of the middle two scores}}{2} = \frac{6 + 7}{2} = 6.5$$

This convention applies even if the scores near the median are not adjacent.

For example: In a series of scores 4, 5, 6, 10, 11, 14
The median is the average of the two middle scores.

$$\text{Median} = \frac{6+10}{2} = 8.$$

(iii) *When there is a duplication of scores near the median:*

When more than one instance of a score value falling near the median exists, the median is obtained by interpolation.

Even Number

Example: Consider the distribution, 4, 5, 6, 6, 6, 7, 7, 8.

The situation is diagrammed below in Figure 3.2. The scores occupying the space on the scale of measurement between their real limits have been shown. Since four of the eight scores are required to be below the median, the median must fall within the interval 5.5—6.5. Since two scores already fall below that interval, two of the three scores existing between 5.5 and 6.5 are required to be below the median. Therefore, two-thirds or .67 of the one unit interval is added to its lower limit: $5.5 + .67 = 6.17$. Examine Figure 3.2 carefully and consider the fractions of frequencies and you will notice that the distribution is divided in half at the point 6.17.

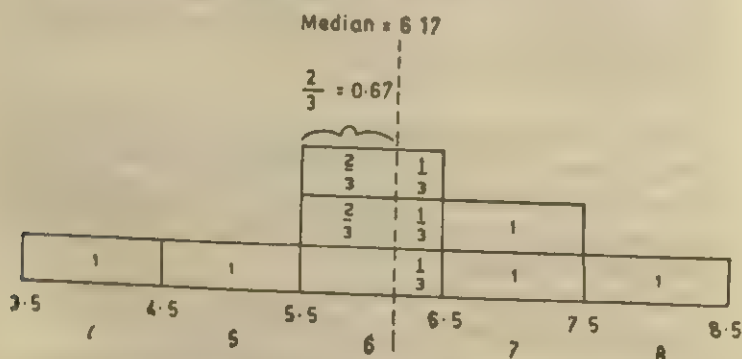


Fig. 3.2. Computation of the median when there is duplication of scores (Even number).

Odd Number

Example: When there are an odd number of scores, the solution proceeds in the same manner as above.

Suppose a distribution contains the following nine scores: 4, 5, 6, 6, 6, 7, 7, 8, 8. The median falls within the score interval 5.5 to 6.5. In the diagram shown below (Fig. 3.3) the scores occupying the space on the scale of measurement between their real limits have been shown. Since there are, in all, 9 scores in the distribution, $4\frac{1}{2}$ of them must be below the median. Again, two scores exist below 5.5 and therefore $2\frac{1}{2}$ of the three scores in one interval 5.5–6.5 must fall below the median.

That is $\frac{2\frac{1}{2}}{3} = \frac{5/2}{3} = \frac{5}{6} = .83$ of the one-unit interval

must fall below the median. Therefore,

$$\text{the median} = 5.5 + .83 = 6.33$$

The student may notice that in Fig. 3.3, the point 6.33 divides the distribution into exact two halves.

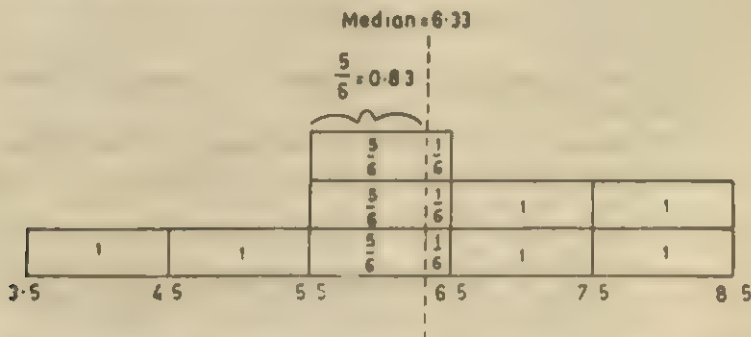


Fig. 3.3. Computation of Median when there is Duplication of Scores (Odd Number).

The procedure and the logic presented above is quite simple and follows a layman's approach to such problems. We can also have a formula for the calculation of median in such situations:

Formula

$$Md = L + \left[\frac{N/2 - F_b}{f_w} \right] i \quad (3.8)$$

In which *symbols**Values in the
above example*

L = The lower limit of the interval containing the Md	5.5
N = The number of scores in the total distribution	9
F_b = The number of cases falling below the lower limit of interval containing the median	2
f_w = The number of scores within the interval containing the median	3
i = the size of the class interval.	1

Substituting the values in the formula, we have

$$\begin{aligned}
 Md &= 5.5 + \left[\frac{9/2 - 2}{3} \right] \times 1 \\
 &= 5.5 + \frac{5/2}{3} \\
 &= 5.5 + .83 \\
 &= 6.3
 \end{aligned}$$

3.2.2 Calculation of Median from a Frequency Distribution

In calculating the median from data grouped in the form of a frequency distribution, the problem is to determine a value

TABLE 3.7

Calculation of Median

(1) <i>Class interval</i>	(2) <i>Exact limits</i>	(3) <i>Frequency</i>	(4) <i>Cumulative frequency</i>
45—49	44.5—49.5	2	50
40—44	39.5—44.5	3	48
35—39	34.5—39.5	2	45
30—34	29.5—34.5	6	43
25—29	24.5—29.5	8	37
20—24	19.5—24.5	8	29 Md. lies in this CI

15—19	14.5—19.5	7	21
10—14	9.5—14.5	5	14
5—9	4.5—9.5	9	9

$$N=50$$

of the variable such that one half the observations fall above this and the other half below. The fundamental logic of calculation remains the same as already described above in relation to the ungrouped data. The method will be illustrated with reference to the data in Table 3.7.

Formula

$$\text{Median} = L + \left[\frac{N/2 - F_b}{f_w} \right] \times i$$

where, L = exact lower limit of the CI in which Median lies

F_b = cumulative frequency below the CI containing Median

f_w = frequency within the CI containing Median

i = size of class interval.

Here, $L = 19.5$

$F = 21$; $f_w = 8$; $i = 5$.

$$\begin{aligned} \text{Median} &= 19.5 + \left[\frac{50/2 - 21}{8} \right] \times 5 \\ &= 19.5 + \left[\frac{25 - 21}{8} \right] \times 5 \\ &= 19.5 + (4/8) \times 5 = 19.5 + 2.5 = 22. \end{aligned}$$

Computational Steps

First, record the cumulative frequencies as shown in Column 4.

Second, determine $N/2$, one half the number of cases, in this example, $50/2 = 25$.

Third, identify the class interval in which the 25th case, the middle case, falls. In this example, it is in class interval 20—24 with exact limits 19.5—24.5.

Fourth, interpolate, between the exact limits of the interval to find a value above and below which 25 cases lie. Observe that 8 cases fall within the limits 19.5—24.5. We assume that these 8 cases are uniformly distributed in rectangular fashion between these exact limits. Now to arrive at the 25th, or middle case we require 4 of the 8 cases within this interval because the cumulative frequency below this interval is 21 which shows that 21 cases have been covered up to 19.5, the upper limit of CI 15—19. This means that we find a point between 19.5 and 24.5 such as 4 cases fall below and 4 cases fall above it. The proportion of the interval we require is $4/8$ which is $4/8 \times 5$ units of scores, or 2.5. We add this to the lower limit of the interval to obtain median, which is $19.5 + 2.5 = 22.00$.

The formula and the calculations shown in Table 3.7 are quite easy and the student can follow the same with convenience. However, the steps mentioned above are summarised below:

1. Compute the cumulative frequencies.
2. Determine $N/2$ or one-half of the cases.
3. Find the class interval in which the middle case falls and determine the exact limits of the interval.
4. Interpolate to find a value on the scale below which and above which one-half of the total number of cases falls. This is the median.

3.2.3 Calculation of Median when the Frequency Distribution contains Gaps

Students may experience difficulty in the calculation of Median when there are gaps or zero frequency upon one or more intervals near the centre of the distribution. The method to be followed in such cases is shown in Table 3.8 below.

TABLE 3.8

Computation of the Median from Distribution with Gaps

<i>Class Intervals (scores)</i>	<i>f</i>
35-39	3
30-34	5
25-29	2
20-24	0
15-19	0
10-14	2
5-9	4
0-4	4
	<hr/>
	N=20
	N/2=10

$$\begin{aligned}
 \text{Mdn} &= L_r + \frac{N/2 - \text{Cum } f_b}{f_w} \times i \\
 &= 9.5 + \frac{10 - 8}{2} \times 10 \\
 &= 9.5 + 10 = 19.5
 \end{aligned}$$

Value of i in this case is 10, the size of the extended class interval of 10—19.

Since $N=20$, $N/2=10$, count up the frequency column 10 scores from below. Ordinarily 14.5, the upper limit of CI 10-14 should have been taken as the median. However, by counting down 10 scores from the frequency Col., we arrive at 24.5, the lower limit of CI 25—29. To resolve this discrepancy in the value of the median by the two approaches of counting up from below and counting down from above, we extend the middle class intervals. Here, the CI 10—14 is extended upto 19 with a new size of 10; the CI 25—29 is extended down to 20, with a new size of 10. Lengthening of these intervals removes the zero frequency on the adjacent intervals by spreading the numerical frequency over to CI's having zero frequencies and creating confusion in the correct calculation of the median. Now

counting up from below, we complete 10 frequencies at 19.5 the upper limit of CI 10–19. Counting down from above also gives a median of 19.5, the lower limit of CI 20–29. Computation from the two ends of the distribution with extended CI's near the median, as shown in Table 3.8 now gives consistent results.

In cases where there is only one zero frequency exactly at the centre of the distribution, the midpoint of the interval having this zero will be the median. The same logic can be used in cases having three or more zeroes.

3.3. The Mode (Mo)

The *mode* is the most frequently occurring score. When a frequency distribution is used, the mode is the midpoint of the interval with the largest number of cases or frequencies. If two adjacent scores have the same frequency and the frequencies are the highest in the distribution, then the mode is the sum of the two scores divided by two. When there are two non-adjacent scores with the same frequency and they are the highest in the distribution, each score may be referred to as the "mode" and the distribution is *bimodal*. Consider the following sets of scores to understand the method of calculation of mode from each.

Set I

Scores: 5, 5, 10, 10, 10, 11, 11, 11, 13, 13, 13, 13, 14, 14, 15, 15.

Mode = Since score 13 occurs the largest number of times ($f=4$), it is the value of the mode.

Set II

Scores: 5, 5, 10, 10, 12, 12, 12, 13, 13, 13, 14, 14, 15, 15.

Mode = The adjacent scores of 12 and 13 have the largest but equal frequency of 3 each, hence the average of these two values will be the mode which is $(12+13)/2=12.5$.

Set III

Scores: 5, 5, 10, 10, 12, 12, 12, 13, 13, 14, 14, 14, 15, 15.

Mode The non-adjacent values of 12 and 14 have the largest but equal frequency of three each. Hence, this set of scores has two modes, 12 and 14. It can thus be called as *bimodal*.

Set IV

Scores: 7, 7, 8, 8, 10, 10, 11, 11, 13, 13, 16, 16.

Mode Here all values occur with a frequency of 2, hence do not permit the calculation of a modal value. In this case, mode is indeterminate. It is a rectangular shaped distribution with equal frequency on all the score points.

3.3.1 Calculation of Mode in a Frequency Distribution

When scores are grouped in frequency distribution, the mode is the midpoint of the interval with the largest frequency. Consider the following frequency distribution.

Scores	<i>f</i>	
35—39	3	
30—34	4	
25—29	6	Mode is the midpoint of class interval 15—19 which has the largest frequency of 9. Hence $\text{mode} = \frac{\text{Lower limit} + \text{Upper limit}}{2} \quad (3.9)$ $= \frac{15 + 19}{2} = 17.$
20—24	7	
15—19	9	
10—14	6	
5—9	5	
	40	

This is also called the *crude mode* or the *empirical mode*. It may be distinguished from the *true mode* which is the point (or peak) of the greatest concentration of scores in the distribution. The crude mode is approximately equal to the true mode and serves most of the practical purposes. A formula for approximating the true mode from the symmetrical or not badly skewed distributions is

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean} \quad (3.10)$$

In the frequency distribution given above,

Median=19.5; and Mean=20.4

Hence, Mode= $(3 \times 19.5) - (2 \times 20.4)$
 $=17.7$

The mode is a measure of very limited practical value. It does not lend itself readily to further algebraic manipulations. It does not, in general, enter into the calculations of further statistical measures. It may acquire meaning if the number of measurements under consideration is fairly very large.

3 4 Comparison of the Mean, Median and Mode

The essential difference between the mean and the median is that the mean reflects the values of each score in the distribution whereas the median is based largely on the score where the midpoint of the distribution falls without regard for the particular value of many of the scores. For example, consider the following illustration:

<i>Scores</i>	<i>Mean</i>	<i>Median</i>
2, 3, 4, 5, 6	4	4
2, 3, 4, 5, 36	10	4
2, 3, 4, 5, 76	16	4

In the above table, only the last number differs from one distribution to the other. The mean reflects these differences but the median does not. The median is the midpoint of the distribution and has an equal number of cases falling on both sides of it. The value of the extreme scores does not matter but only the fact of their existence is taken into consideration. The numerator of the formula for M , ΣX shows that each score is to be summed up. Thus changing a score value will change the value of the mean.

The mode is a simple measure of central tendency and reflects only the most frequently occurring score. Its use is restricted to a very few problems in Social Sciences.

The mean, median and mode are sensitive to different aspects of a group of scores, generally they are not the same in a given distribution. Their relative positions can be seen from Figure 3.4.

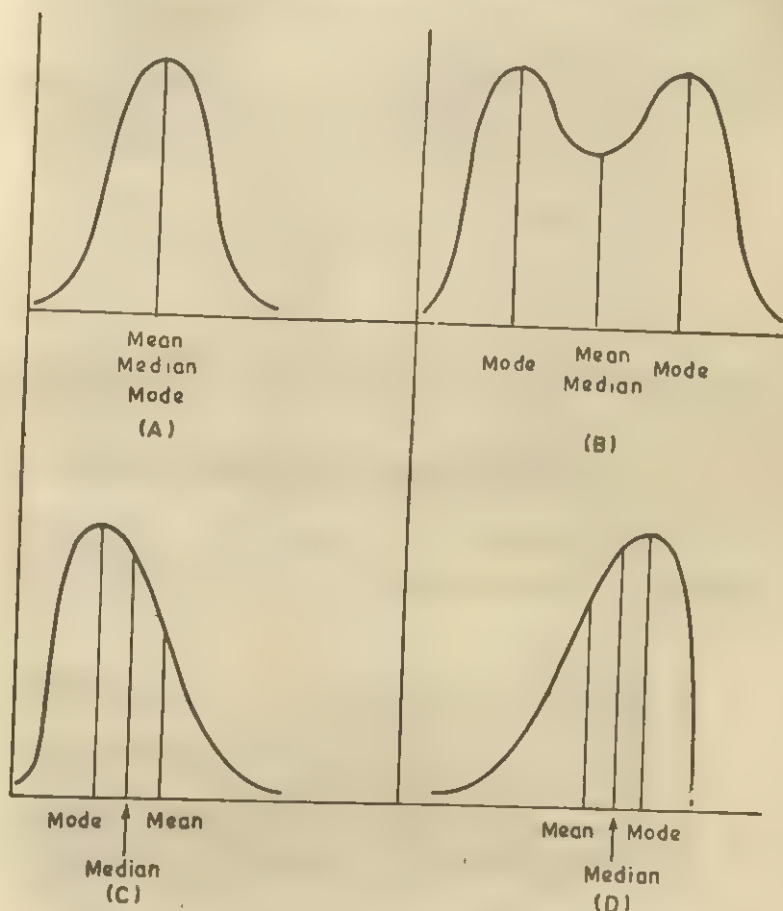


Fig. 3.4. The Relative Position of Mean, Median and Mode in Different Types of Distributions:

- (A) Symmetrical Unimodal; (B) Symmetrical Bimodal;
(C) Positively Skewed; and (D) Negatively Skewed.

- (i) If the distribution is symmetrical and unimodal (having one mode only), the mean, mode and median fall at the

same point i.e., they have the same value (See Part A of Figure 3.4).

- (ii) If the distribution is symmetrical but *bimodal* (having two modes), the mean and the median fall at the same point but the values of the modes are different (See Part B of Figure 3.4).
- (iii) If the scores are bunched on the lower side of the mean (*Positively skewed distribution*), the value of the mean is higher than that of the median (See Part C of Figure 3.4).
- (iv) If the scores are bunched on the upper side of the mean (*negatively skewed distribution*), the value of the median is higher than that of the mean (See Part D of Figure 3.4).

Thus Part C and D of Figure 3.4 show that the mean is always pulled more towards the skewed end of the distribution than is the median.

3.5 Guidelines for the Use of Various Measures of Central Tendency

The following simple rules may prove helpful to the puzzled student as to when to use the various measures of central tendency.

Mean is useful

1. When scores are symmetrically or nearly symmetrically distributed around a central point.
2. When the situation warrants a measure of central tendency, having the greatest stability.
3. When the researcher wishes to compute SD, coefficient of correlation and other statistics which are based upon the mean.

Median is useful

1. When the exact midpoint separating the distribution into two equal halves is wanted.
2. When extreme scores are present in the distribution. Extreme scores affect the mean more than the median.
3. When the number of scores above or below the central tendency is known but not their exact values.

4. When the upper exact limit of the top class interval (or the lower exact limit of the lowest class interval) is not known, i.e., the frequency distribution is not complete.

Mode is useful

1. When a quick measure is all what is wanted.
2. When an approximate measure of central tendency would do.
3. When only the most typical value is required. For example, the most typical size of the shirt or shoes worn by an average man.

Exercises for practice

- 3.1 Calculate Mean, Median and Mode from the following ungrouped scores.

- (a) 2, 8, 7, 10, 5, 2, 1
 (b) 20, 15, 14, 14, 16, 13, 12
 (c) 10, 12, 13, 15, 14, 15

- 3.2 Calculate mean, median and mode from the following frequency distributions.

(a)		(b)		(c)	
Class interval	f	Class interval	f	Class interval	f
80—89	6	190—199	2	45—49	2
70—79	9	180—189	4	40—44	3
60—69	10	170—179	15	35—39	6
50—59	25	160—169	11	30—34	9
40—49	25	150—159	38	25—29	13
30—39	10	140—149	8	20—24	8
20—29	5	130—139	6	15—19	6
10—19	7	120—129	2	10—14	2
0—9	3	110—119	7	5—9	1
		100—109	7		
— — —	— — —	— — —	— — —	— — —	— — —
	N=100		N=100		N=50

- (d) Calculate mean from the three distributions by long method.
- 3.3 Define mean, median and mode.
- 3.4 When should mean, median and mode be calculated?
- 3.5 What are the relative merits of mean and median?
- 3.6 Give examples from day-to-day life when we make use of mode without being aware of it.

CHAPTER 4

MEASURES OF VARIABILITY

Measures of central tendency summarize only one special aspect of a distribution. Any distribution has at least one more feature that must be summarized in some way. Distributions exhibit spread or dispersion, that tendency for observations to depart from Central tendency. Variability or dispersion is thus an important concept in statistical inquiry. It reflects the "poorness" of central tendency as a description of a randomly selected case as it depicts the tendency of observations not to be like the average. As variability is accounted for, estimates and inferences are improved.

Look at the following sets of scores and try to visualize the correctness of the statement made above.

TABLE 4.1
Three sets of scores with Equal Means but
Different Dispersions

<i>Sr. No.</i>	<i>Set 1</i>	<i>Set 2</i>	<i>Set 3</i>
1.	10	13	19
2.	10	12	16
3.	10	11	13
4.	10	10	10
5.	10	9	7
6.	10	8	4
7.	10	7	1
Mean	10	10	10

All the three sets have means equal to 10. The variability in Set 1 is zero as each of the seven scores equals the mean. The dispersion or spread of scores in Set 3 is greater than in Set 2. The distances of scores (deviations) from the mean are larger in Set 3 than in Set 2. A description of scores of Set 1 on the basis of central tendency mean is free from error while this description from Set 2 and Set 3 involves error which is larger in the case of Set 3.

The following measures of dispersion or variability will be discussed in this chapter. Each of these provides a numerical index of the variability of the scores.

1. The Range
2. The Mean Deviation or Average Deviation (AD)
3. Variance
4. Standard Deviation
5. Semi-Interquartile Range

4.1 The Range

The Range or the total range is the distance given by the highest score minus lowest score in the distribution.

$$\text{Range} = X_{\text{Max}} - X_{\text{Min}} \quad (4.1)$$

The ranges in the three sets of scores given in Table 1 can be calculated as follows:

TABLE 4.2
Calculation of Range

Set	Highest score		Smallest score		Range
1	5	Minus	5	=	0
2	13	Minus	7	=	6
3	19	Minus	1	=	18

Interpretation: Set 1 : All scores are covered within a score distance of zero units.

Set 2 : All scores are covered within a score distance of six units.

Set 3 : The student may interpret this result himself.

Technically, the range should probably be defined as the difference between the upper real limit of the largest score minus the lower real limit of the smallest score. Since the range is at best our approximate index of variability, it does not seem appropriate to insist upon this level of accuracy.* As evident from Formula I, range takes into account the extremes of the scores only and ignores others. Hence it suffers from the following limitations:

Limitations

1. It is unreliable when N is small or when there are gaps (i.e. zero f 's) in the frequency distribution.
2. A change in the value of either of the highest score or the lowest score leads to a change in its values.
3. It does not consider the value of the scores between the highest and the smallest scores and does not reflect the change if made in them.
4. Further statistical analysis are difficult to make.

Uses

However range can be used with profit in the following situations:

1. When a quick and crude estimate of variability is all what is desired.
2. When data are too scant or too scattered and a more precise measure of variability is not warranted.
3. When the knowledge of only the extreme scores or of total spread is required.
4. When the phenomenon is prone to wide fluctuations such as range of fluctuating temperature of a patient, the daily fluctuating values of a stock and the annual range of temperature values for a particular geographical region.
5. When ease of computation is an important consideration.

4.2 The Average Deviation (AD)

The average deviation is the average distance between the mean and the scores in the distribution. It is the arithmetic

*If this accuracy is insisted upon the values of ranges in the three sets would be: Set 1: $5.5 - 4.5 = 1$, Set 2: $13.5 - 6.5 = 7$; Set 3: $19.5 - 0.5 = 19$.

mean of all the deviations when algebraic signs are disregarded. The deviation is defined as the distance of the score from the mean of the distribution. Scores larger than the mean will have positive or plus signs, and those, smaller than the mean, negative or minus signs. Scores which coincide with the mean will have zero deviation. Algebraically, deviation can be defined, $x = X - M$ (A deviation of a score from the mean) (4.2)

where X = original score; and M = arithmetic mean.

The sum of the deviations (with algebraic signs) from the arithmetic mean is always zero. (See chapter on Measures of central tendency). Hence their average will also be zero and thus, useless for measuring and describing dispersion. Hence statisticians decided to disregard the algebraic signs and the direction of the deviations. Only the sizes of the deviations are taken into account. The formula for the calculation of average deviation is:

$$AD = \frac{\sum |x|}{N} \text{ (The average deviation)} \quad (4.3)$$

where Σ = sum of; $|x|$ = absolute value of deviation; and N = Total number of scores or observations.

TABLE 4.3

**Calculation of Average Deviation (AD) from
Set 3 of Table 4.1**

Persons Scores Deviation Deviation

	<i>X</i>	<i>with sign</i> <i>x = X - M</i>	<i>without</i> <i>sign</i> <i> x </i>
1.	13	+3	3
2.	12	+2	2
3.	11	+1	1
4.	10	0	0
5.	9	-2	1
6.	8	-3	3
Σ	60	0	12

$$M = 10$$

$$\Sigma x = 12$$

$$N = 7$$

$$AD = \frac{12}{7} = 1.71$$

Interpretation: The result means that the scores deviated, on the average, 1.71 points from the mean.

This technique provides a reasonably stable estimate of variation. It takes into consideration all the scores and the changes that may be incorporated in any one of them. It is easier to calculate as compared to standard deviation. However, it lacks algebraic properties (sign or direction of the deviation is ignored) and cannot be used with other more advanced statistical techniques.

4.3 The Variance and Standard Deviation

A more stable index that reflects the degree of variability in a group of scores is the Variance, and its derivative, the Standard Deviation.

In a previous section, it was shown that in any frequency distribution, the mean deviation from the mean must be zero. Hence the device to get around the difficulty is to take the square of each deviation from the mean, and then to find the average of these squared deviations:

$$\text{Variance, } \sigma^2 = \frac{\Sigma(X - M_X)^2}{N} = \frac{\Sigma x^2}{N} \quad (4.4)$$

Here the symbols are: Σ — Summation; X — Any raw score; M_X — Mean of X scores; N — Number of cases; x^2 — Square of deviation from the mean; σ — Greek letter sigma.

In a grouped distribution, for each interval, the deviation of the midpoint from the mean is squared and multiplied by the frequency for that interval. When this has been done for each interval, the average of these products is the variance. The formula thus becomes:

$$\text{Variance } (\sigma^2) = \frac{\Sigma fx^2}{N} \quad (4.5)$$

Where f stands for the frequency in each class interval; other symbols, as above.

Standard Deviation is derived from the variance by taking the square root of the latter. The formula for the calculation of the standard deviation is thus as follows:

$$\text{SD or } \sigma = \sqrt{\text{Var.}} = \sqrt{\frac{\Sigma x^2}{N}} \quad (4.6)$$

(σ , is pronounced as sig-mah' and used to denote SD.)

Although variance is an adequate way of describing the degree of variability in a distribution, it does have one drawback. The variance is a quantity in squared units of measurement. For example, if measurements are taken in inches, then the mean is some number of inches, and a deviation from the mean is a difference in inches. However, the square of a deviation is a square-inch units, and thus the variance, being a mean squared deviation, must also be in square inches. Thus the problem of obtaining an index of variability in original units arises. This has been taken care of by further calculating the square root of the mean squared deviation. This process converts the index of variation from the square to the linear measure and gives us the root mean square deviation or the standard deviation. Hence Standard deviation is the square root of the variance or square root of the mean squared deviation and is an index of variability in the original units.

In the calculation of SD, deviations are always taken from the mean, never from the median or mode. The value of S.D. is always positive.

Standard deviation has been termed so because it provides a standard unit for measuring distances of various scores from their mean.

4.3.1 Methods of Calculating Variance and Standard Deviation from Ungrouped Data

The conceptual definitions of the variance and the SD are based on the formulas which incorporate the deviation score method as shown in Formulas 4.4 to 4.6 in this chapter. However to avoid inconvenience of working with fractional values and when a calculating machine is available, the following formulas which are mathematically equivalent to them are also available. In the example solved below the use of both the types of formulas has been made:

Steps in the Computation of SD

(a) Deviation Score Method

- (i) Calculate Mean (ii) Calculate Deviation of each score from the mean.

TABLE 4.4

Calculation of SD from Ungrouped Scores

Deviation Score Method			Raw Score Method	
Score	Deviation from the Mean ($X-M$)	Squared deviation x^2	Score	Squared Score X^2
X	x	x^2	X	X^2
10	+2	4	10	100
7	-1	1	7	49
9	+1	1	9	81
			6	36
6	-2	4	8	64
8	0	0		
		10	40	330

$$\Sigma X = 40$$

$$(\Sigma x^2)$$

$$\Sigma X$$

$$\Sigma X^2$$

$$M = 40/5$$

$$= 8$$

$$N = 5$$

Formula:

$$\text{Var.} = \frac{\Sigma x^2}{N}$$

$$\Sigma x^2 - \frac{(\Sigma x)^2}{N} \quad (4.7)$$

Substituting the values

$$\text{Var.} = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}$$

$$\text{Var.} = \frac{10}{5} = 2$$

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} \quad (4.8)$$

$$\sigma = \sqrt{2} = 1.414$$

Substituting the values

$$\text{Var.} = \frac{330 - \frac{(40)^2}{5}}{5}$$

$$= \frac{10}{5} = 2$$

$$\sigma = \sqrt{2} = 1.414$$

Interpretation: The scores, on the average, vary or deviate 1.414 units from their mean.

- (iii) Square each deviation,
- (iv) Sum up the squared deviations to obtain Σx^2 .
- (v) Substitute the values in the formula and solve.

(b) *Raw Score Method*

- (i) List up the scores, X
- (ii) Sum up the scores to obtain ΣX .
- (iii) Square each score.
- (iv) Sum up the squared scores.
- (v) Substitute the values in the formula and solve.

4.3.2 Calculation of SD from the Grouped data

When scores are arranged in the form of score bands or class intervals and frequencies are shown against each class interval calculation of SD can be undertaken by using a long method or a short method. Both of these methods are demonstrated below with the help of solved examples.

Long Method

TABLE 4.5

Calculation of SD by Long Method (Using Real Mean)

Class interval	Midpoint X	Frequency f	Deviation of (X) from Mean x	fx	fx^2
(1)	(2)	(3)	(4)	(5)	(6)
45-49	47	2	24.6	49.2	1210.32
40-44	42	3	19.6	58.8	1152.48
35-39	37	2	14.6	29.2	426.23
30-34	32	6	9.6	57.6	552.96
25-29	27	8	4.6	36.8	169.28

(1)	(2)	(3)	(4)	(5)	(6)
20-24	22	8	— .40	—3.2	1.28
15-19	17	7	—5.40	—37.8	204.12
10-14	12	5	—10.40	—52.0	540.80
5-9	7	9	—15.40	—138.6	2134.44
Mean=22.40		N=50		6392.00	
				Σfx^2	
$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{6392.00}{50}} = \sqrt{127.84} = 11.31$					

Steps in the Computation of SD by Long Method

- (i) Write down the class intervals and frequencies as shown in Cols (1 and 3),
- (ii) Find out the midpoint of each class interval,

$$\text{Midpoint} = \frac{\text{upper Limit} + \text{Lower Limit}}{2}$$
- (iii) Calculate Mean by using any method described in a previous chapter.
- (iv) Obtain deviations from Mean, (Mid-point—Mean) as in Col. (4)
- (v) Multiply f and x , [Col. (3) \times Col. (4)] to obtain fx .
- (vi) Multiply fx and x i.e. Col. (4) and (5) to obtain fx^2 as in Col. (6)
- (vii) Sum up Col. (6) to obtain Σfx^2
- (viii) Substitute the values in the formula and solve.

Short Method

When large values of scores are involved, it is better to use the short method for calculating SD. In this method, like the calculation of Mean by the Assumed Mean method, deviations are taken from the assumed mean. The detailed procedure is given below:

TABLE 4.6

Calculation of SD by short Method (Deviations taken from Assumed Mean)

Class interval	Mid-point <i>X</i>	Frequency <i>f</i>	Deviation of (<i>X</i>) from AM in units of CI <i>x'</i>	<i>fx'</i>	<i>fx</i> ²
(1)	(2)	(3)	(4)	(5)	(6)
45-49	47	2	5	10	50
40-44	42	3	4	12	48
35-39	37	2	3	6	18
30-34	32	6	2	12	24
25-29	27	8	1	8	8
20-24	22	8	0	0	0
15-19	17	7	-1	-7	7
10-14	12	5	-2	-10	20
5-9	7	9	-3	-27	81
		50		4	256
		(N)		($\Sigma fx'$)	(Σfx^2)

$$\sigma = i \sqrt{\frac{\Sigma fx'^2 - C^2}{N}} \quad (4.9)$$

in which, *i* stands for the size of class interval; and *C*, for correction.

Here: *i* = 5; $\Sigma fx'^2 = 256$; *N* = 50; and

$$C = \frac{\Sigma fx'}{N}$$

Substituting the values in the formula

$$C^2 = \left(\frac{4}{50} \right)^2 = .0064$$

$$\sigma = 5 \sqrt{\frac{256 - .0064}{50}}$$

$$= 11.31$$

Computational Steps

- (i) Arrange the scores and f 's as in Cols. (1) & (3),
- (ii) Find out the midpoints of all the class intervals and write in Col. (2).
- (iii) Take a midpoint as an assumed mean. This point should be close to the middle of the distribution and as far as possible should have the largest f 's.
- (iv) Deviations (x') are taken from the assumed mean (here, 22) in units of class interval. It can be a mechanical process. Assign 0 to the class interval in which the assumed mean lies. Go on assigning +1, +2, +3, etc. to the class intervals above the mean and -1, -2, -3, etc. to those below the mean. (Col.4)
- (v) Multiply cols. (3) and (4) to obtain fx' and sum up to obtain $\Sigma fx'$
- (vi) Multiply cols. (4) and (5) to obtain fx'^2 and sum up to obtain $\Sigma fx'^2$
- (vii) Find out the value of C which is $-\frac{\Sigma fx'}{N}$
- (viii) Substitute these values in the formula and solve.

Some other formulas for the calculation of SD are given below:

$$\sigma = \sqrt{\frac{\Sigma X^2 - M^2}{N}} \quad (4.10)$$

$$\sigma = \sqrt{\frac{N \Sigma X^2 - (\Sigma X)^2}{N}} \quad (4.11)$$

The symbols are as explained in the previous formulas.

Note: When sample SD is to be used as an estimate of the population SD, the denominator of the formulas will have $N-1$, instead of N , as the former is considered as an unbiased estimate.

4.3.3 Properties and Uses of Variance and SD as Measures of Variability

- (i) The variance is proportional to the average squared deviation of each score from every other score. Hence it indeed reflects the variability of the scores.

- (ii) Since all deviations are squared, the variance will always be positive. The SD is the positive square root of variance and hence will always be positive.
- (iii) If there is no variability among the scores, that is, all the scores in the distribution are identical the value of variance and also of SD will be zero. As variability of scores increases, the variance also increases.
- (iv) Variance and SD are more sensitive to variability in a group of scores and are less variable in themselves.
- (v) The variance and SD are frequently used in other statistical analysis and manipulation and hence are more important than the other measures of variability.
- (vi) The variance can be partitioned into different parts attributed to different sources and hence finds its use in analysis of multivariate factorial designs.
- (vii) Variance and SD are used when statistic of the greatest stability is sought.
- (viii) Variance and SD should be used when extreme deviations are likely to exercise a proportionally greater effect upon the variability.

4.4 The Semi-Inter-Quartile Range or Q

The semi-inter quartile range which is also known as quartile deviation can be defined as half of the difference between the 75th percentile and the 25th percentile. Hence it is one half the scale distance between the 75th and 25th percentiles in a frequency distribution. The 25th percentile is Q_1 or the first quartile on the score scale. The 75th percentile is Q_3 or the third quartile on the score scale. Hence the formula for the calculation of Q is

$$Q = \frac{Q_3 - Q_1}{2} \text{ or } \frac{P_{75} - P_{25}}{2} \quad (4.12)$$

Hence to find out Q, it is essential to calculate the values of Q_3 (or P_{75}) and Q_1 (or P_{25}). Their calculation follows the same procedure as the calculation of median as explained in the previous chapter. These formulas are

$$Q_1 \text{ (or } P_{25}) = L + \frac{(N/4 - \text{Cum}f_b)}{f_{\square}} \times i \quad (4.13)$$

$$Q_3 \text{ (or } P_{75}) = L + \frac{3N/4}{f_q} \text{Cum } f_b \times i \quad (4.14)$$

in which,

L = the exact lower limit of the interval in which the quartile falls.

i = the size of the class interval.

$\text{Cum } f_b$ = Cumulative f below the interval which contains the quartile; f_q = the f on the interval which contains the quartile.

TABLE 4.7

Calculation of Q_1 , Q_3 and Quartile Deviation

(1)	(2)	(3)	
Class interval (X)	Frequency (f)	Cummulative frequency (Cum f)	
45-49	2	50	
40-44	3	48	
35-39	2	45	
30-34	6	43	Q_3 lies in this CI
25-29	8	37	
20-24	8	29	
15-19	7	21	
10-14	5	14	Q_1 lies in this CI
5-9	9	9	
<hr/> N=50 <hr/>			

Calculation of Q_1

Here, $L=9.5$; $\text{Cum. } f_b=9$, $f_q=5$;
 $i=5$, $N=50$

Substituting these values in formula (4.13) we have

$$Q_1 = 9.5 + \frac{(50/4-9)}{5} \times 5 = 9.5 + 3.5 = 13.00$$

Calculation of Q_3

Here, $L = 29.5$; Cum. $f_b = 37$; $f_q = 6$; $i = 5$; $N = 50$.

Substituting these values in formula (4.14) we have

$$\begin{aligned} Q_3 &= 29.50 + \frac{(3 \times 50 - 4 - 37)}{6} \times 5 \\ &= 29.50 + .42 = 29.92 \end{aligned}$$

Calculation of Q

Substituting the values of Q_1 and Q_3 in formula (4.12) we have

$$Q = \frac{29.92 - 13.00}{2} = \frac{16.92}{2} = 8.46$$

Properties and uses of Q

1. In a distribution which is symmetrical around the mean, or when it is normal — Q marks off the 25 per cent cases just above, and the 25 per cent cases just below the median.
2. It is a measure of the variability of the middle 50 per cent cases and ignores the 25 per cent cases in each of the two tails.
3. It should be used when a measure of dispersion of the concentration of 50 per cent of the cases around the median is required.
4. It should be used when the measure of central tendency to be used is the median.
5. It is a better measure, when there are scattered or extreme scores which would influence the SD disproportionately.
6. Q is known as the Probable Error (PE) in a normal distribution.

In this chapter, some important and more popular measures of variability or dispersion have been presented. These indices show the 'spread' or 'scatter' of the separate scores around their central tendency. One of these should be reported along with the relevant measure of central tendency to provide a better description of the distribution.

4.5 Relationship Between Sum of Squares, Variance and SD

The concepts of sums of squares ($\sum x^2$), variance (V), and Standard Deviation (SD) are very closely related with each other. These concepts are also very important in all statistical work. Hence, this section is devoted to explain them.

In verbal terms, a standard deviation is the square root of the arithmetic mean of the squared deviations of scores from their mean. It can thus rightly be termed as *root-mean-square deviation*. Consider Table 4.8 shown on p. 75. In column (3) we have the deviations of each score from the mean. The sum of these deviations is always zero. In column (4), we have the squares of these deviations (x^2). The sum of these squared deviations ($\sum x^2$) is briefly called as *sum of squares*, which is equal to 88 in this case. Variance is the mean of the sum of squares, which is equal to $88/7 = 12.57$, the SD is the square root of variance which in this case is,

$$SD = \sqrt{12.57} = 3.55$$

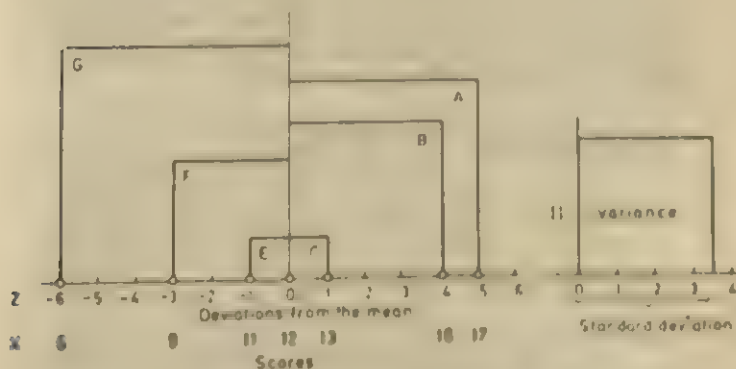


Fig. 4.1 Comparison of Standard Deviation and Variance

(Guilford, 1973)

A geometrical representation of the above ideas is shown, as usual, in the form of a straight line extending from left to right. The original score values have also been marked below

the relevant points. The mean score of 12 has been shown as deviation equal to zero and taken as the main reference point. All seven persons have been shown to retain their relative positions, in correct rank order and at the same separations as in original scores.

The deviations have been represented by linear distances. Hence, the squared deviations as shown in Col. 4, have been represented in terms of *areas* namely squares. The squares belonging to different individuals A to G are shown in Figure 4.1(I).

The sum of squares would be represented geometrically as an area equal to a composite of all the squares in Figure 4.1(I). This could also be shown as a square or as a rectangle. Its surface will have 88 units each unit equal in size to those representing persons C and E. The apportioning of the total area ($\sum x^2$) equally among the seven individuals amounts to taking the arithmetic mean of it. This is variance which is represented by the square in Figure 4.1 (II). The baseline of this square has units equal to those in the larger diagram. The length of its side is the square root of its area and represents the standard deviation. The inter-relationships of sum of squares, variance and standard deviation can also be shown algebraically as follows:

$$\text{Standard deviation, } \sigma = \sqrt{\frac{\sum x^2}{N}} = \sqrt{V}$$

$$\text{Variance, } V = \frac{\sum x^2}{N} = \sigma^2$$

$$\text{Sum of squares, } \sum x^2 = NV = N\sigma^2$$

Both V and σ are indicators of amount of variability or dispersion in a distribution. V or σ^2 is said to measure variance and σ , to measure variability. V and σ also form familiar indicators of the extent of individual differences which form the basis of all psychological and educational testing.

TABLE 4.8

Data Illustrating Sum of Squares (Σx^2), Variance (V) and Standard Deviation (SD)

<i>Person</i>	<i>Score</i> X	<i>Deviations</i> x	<i>Deviations squared</i> x^2
(1)	(2)	(3)	(4)
A	17	+5	25
B	16	+4	16
C	13	+1	1
D	12	0	0
E	11	-1	1
F	9	-3	9
G	6	-6	36
Sum	84	0 Σx	88 Σx^2
Mean	12.0	0.0	12.57 = V
Standard deviation			3.55 = σ

Exercises for practice

1. Calculate variance and SD from the following distribution of scores (use long method).

<i>Class Interval</i>	<i>f</i>
190-199	2
180-189	4
170-179	15
160-169	11
150-159	38
140-149	8
130-139	6
120-129	2
110-119	7
100-109	7

N = 100

6. The results of the investigation of the
the following questions are as follows:
7. The results of the investigation of the
the following questions are as follows:
8. The results of the investigation of the
the following questions are as follows:
9. The results of the investigation of the
the following questions are as follows:
10. The results of the investigation of the
the following questions are as follows:

SEE US ON OUR STAND.

The proposed Board of Directors is composed of seven members, three of whom are representatives of the Government of the District of Columbia. The Board will be responsible for the management and operation of the Corporation, and will have the authority to issue bonds, to borrow money, and to enter into contracts.

[illegible]

- (ii) *Units of uniform size*, so that a gain of 5 points on one part of the scale signifies the same thing as a gain of 5 points on any part of the scale.
- (iii) *A true zero point* of "just none of" the quality in question, so that we can legitimately think of scores as representing twice "as much as" or "two-thirds" as much as.

TABLE 5.1

Main Types of Norms for Educational and Psychological Tests

<i>Type of Norm</i>	<i>Type of Comparison</i>	<i>Type of Group</i>
Age Norms	Individual matched to group whose performance he equals.	Successive age groups
Grade Norms	—Same as above—	Successive grade groups
Percentile Norms	Percent of group surpassed by individual.	Single age or grade group to which the individual belongs
Standard Score Norms	Number of standard deviations individual falls above or below average of group.	—Same as above—

The different types of norms developed for tests represent a marked progress toward the first two of the objectives. The third can perhaps be never achieved for the traits with which

psychologists, educationists and other social scientists are concerned. It is more or less impossible to arrive at "a total absence point" of intelligence or arithmetic ability or "sociability" although we may allot a student a zero score on any test. It would merely be an arbitrary zero. The fact of a general prevalence of inequality of units and absence of a real zero point brings us to the conclusion that a raw point score can be given meaning only by referring it to some type of group or groups. Hence, the need for norms.

5.1 Age Norms

The age norm for a particular age is the average value of the trait for persons of that particular age. Age norms can be established for a trait that shows a progressive change with age for example, weight, height, etc. We may take a representative sample of 8-year-old girls and measure their heights. The average value thus obtained will be the age norm for the 8-year-old girls. Similarly, age norms can be set up for the 9-year-olds, 10-year-olds and so on. Later on, each girl's height can be interpreted by comparing it with the average heights and identifying the age groups with whose height it was the closest. *If an 8-year-old girl has a height equal to the norm for the 12-year-olds, we declare that she is as tall as an average girl of 12 years.*

The norms based on age framework are relatively simple and familiar. They are convenient for a trait that shows continuous and relatively steady growth over a period of years. However, age norms suffer from certain disadvantages such as lack of standard and uniform units of growth in height from one year to another. Growth in a trait may slow down or even stop after a particular period. Moreover, these are less appropriate for the non-biological framework of years of growth which may be based on training, schooling or amount of interaction with others.

5.2 Grade Norms

A test is given to representative groups in each of a series of school grades (classes) and the average score determined for each grade (like 5th, 6th, 7th and 8th classes). These averages then represent the norms for various grades. Scores

lying between the norm for two successive grades are assigned fractional credits by interpolation. Generally, the grade value of 5.0 is assigned to average performance at the beginning of the fifth grade, 5.5 to average performance at the middle of the grade, and so forth. The interpretation is similar to that for the age norms. If a child of 6th grade obtains a score equal to the norm for the 8th grade on a test of arithmetic, we say that the child though belonging to 6th grade is yet performing at a level equal to an average child of 8th grade.

Grade norms are relatively easy to determine as the administrative groups based on grades are easily available. They are easy to set up and convenient to interpret. These are more useful for interpreting the academic accomplishments of children especially in primary schools. In the upper grades, slowing down of the progress makes the grade norms less meaningful and inappropriate. These should not be mistaken as the top achievement in the subject or mastery of the subject. They represent the average achievement, neither more nor less.

5.3 Percentiles

The n th percentile is that scale value or score point below which n per cent of the cases in the distribution fall. The scale value of the variable is called a *percentile point* or simply *percentile*, while its corresponding percentage value is known as its *percentile rank*.

While studying the concept and the method of calculation of the median, in the chapter on measures of central tendency, it was noted that median was another similar concept which stood for a score point below which 50 per cent of the cases lie and hence equivalent to P_{50} . Q_1 and Q_3 , introduced in yet another chapter stood for 25th percentile, and 75th percentile respectively indicating respectively the score points below which 25 per cent and 75 per cent cases would lie. The method of calculation of percentiles is thus similar to that of calculating median, Q_1 , or Q_3 .

5.3.1 Calculation of Percentiles from Ungrouped Data:

TABLE 5.2

Computation of Percentile Points from Ungrouped Data (Scores of 40 students on an arithmetic test)

Student Score		Student Score	
1	26	21	54
2	28	22	55
3	28	23	55
4	29 $P_{10} = \frac{29+31}{2} = 30$	24	56
5	31	25	58
6	35	26	58
7	35	27	58
8	35	28	58
9	35	29	59
10	36	30	60 $P_{75} = \frac{60+62}{2} = 61$
11	38 $P_{25} = \frac{36+38}{2} = 37$	31	62
12	39	32	63
13	40 $P_{30} = \frac{39+40}{2} = 39.5$	33	64 $P_{80} = \frac{63+64}{2} = 63.5$
14	41	34	65
15	45	35	65
16	50	36	68
17	51	37	71
18	52	38	72 $P_{95} = \frac{72+73}{2} = 72.5$
19	52	39	73
20	52	40	77

5.3.1.1 When no duplication near Percentile exists

Suppose we desire to calculate P_{30} or the score value below which 30 per cent of the 40 cases lie. Thirty per cent of 40 students $= 40 \times \frac{30}{100} = 12$ students. Therefore, 12 students must

be below P_{30} . The scores in the table have been arranged in an ascending order. Counting up, we find that the 12th student scored 39 and the 13th student scored 40. The desired point must fall between these two score values. It is a convention that the average of such two scores will be taken as the desired percentile. Hence.

$$P_{30} = \frac{39+40}{2} = 39.5$$

In the same manner, the values of the following percentile points have been calculated:

Percentile	Number of cases required below Percentile	Percentile value
P_{10}	$40 \times \frac{10}{100} = 4$	$\frac{4\text{th score} + 5\text{th score}}{2}$ $= \frac{29+31}{2} = 30$
P_{25}	$40 \times \frac{25}{100} = 10$	$\frac{10\text{th score} + 11\text{th score}}{2}$ $= \frac{36+38}{2} = 37$
P_{30}	$40 \times \frac{30}{100} = 12$	$\frac{12\text{th score} + 13\text{th score}}{2}$ $= \frac{39+40}{2} = 39.5$
P_{75}	$40 \times \frac{75}{100} = 30$	$\frac{30\text{th score} + 31\text{st score}}{2}$ $= \frac{60+62}{2} = 61$
P_{80}	$40 \times \frac{80}{100} = 32$	$\frac{32\text{nd score} + 33\text{rd score}}{2}$ $= \frac{63+64}{2} = 63.5$
P_{95}	$40 \times \frac{95}{100} = 38$	$\frac{38\text{th score} + 39\text{th score}}{2}$ $= \frac{72+73}{2} = 72.5$

The student may try the calculation of other percentile points where no duplication of scores occurs at the percentile.

5.3.1.2 When duplication near percentile exists:

The procedure of calculating the value of percentile points when the score value containing the percentile point has more than one frequency is different from the procedure mentioned above.

Example 1

For example, consider the score point corresponding to P_{20} . Twenty per cent of 40 students is 8 students. However, the 8th and 9th, and also the 6th and 7th, students have a score of 35 each. The real lower limit of score of 35 is 34.5. Five cases are covered upto 34.5. To reach the target of 8 cases, 3 more cases are to be taken from the 4 cases which are within the score interval 34.5-35.5 (real lower and upper limits of the score 35) whose size is equal to one score unit. Hence, 3 out of 4 students scoring 35 fall below the 20th percentile. It would mean an addition of $3/4$ th of the class interval (i.e. .75) to the lower limit of 34.5.

$$\text{Hence } P_{20} = 34.5 + .75 = 35.25.$$

This procedure is very much similar to the one already described in a previous chapter on the calculation of median.

Similarly, the following percentile values can also be calculated.

Example 2

$P_{65} = 40 \times \frac{65}{100} = 26\text{th score}$; the class interval of score 58 (exact limits 47.5-58.5) contains 4 students (serial 25 to 28). To reach 26th student, two out of four cases are to be covered over and above the score of 57.5. These two cases span $2/4$ or .5 or half of the CI size. Hence.

$$P_{65} = 57.5 + .5 = 58.00.$$

The student may check the following values also

$$P_{15} = 34.75; P_{45} = 51.83; P_{85} = 65.00.$$

5.4 Calculation of Percentile Ranks From Ungrouped Data

Percentile points answer the question "what is the score point below which a particular given percentage of cases lie?" In such cases, a score point was computed. However, if the question is reversed as follows, "What percentage of cases lie below a given score point?", we have to find out the Percentile Rank (PR) of the score of the person holding that score.

5.4.1 When no Duplication Near Percentile Exists

From Table 5.2, one may ask "What is the percentile rank (PR) of a score of 63?" The score of 63 holds 32nd position in the descending order of the score distribution. This means that 31 persons or $\frac{31}{40} \times 100 = 77.5$ per cent students scored less than the lower limit of 63 which is 62.5. By the same argument, 8 persons or $\frac{8}{40} \times 100 = 20$ per cent scored above the upper limit of the score of 63 which is 63.5. The total percentage of cases below and above the score interval of $62.5 - 63.5 = 77.5$ per cent + 20 per cent = 97.5 per cent. By subtraction, we may find the percentage of cases within the score interval of $62.5 - 63.5$, which is 100.00 per cent - 97.5 per cent = 2.5 per cent (This checks with the fact that only 1 case out of 40 or 2.5 per cent cases are contained in this score interval). Since the midpoint of the class interval, is considered to be the most representative score, we need to divide the class interval into two equal halves to reach a score of 63.00 from 62.5. Hence the percentage of cases contained in this interval is also to be divided into two equal halves; one half of this is then to be added to the percentage of 77.5 covered upto the score of 62.5.

Hence PR of score of 73 = 77.5 per cent + $\frac{2.5}{2}$ per cent = 78.75 per cent.

The percentile ranks are conventionally reported in whole percentages, the percentage of 78.75 is to be rounded off to the closest one i.e. 79 per cent. Hence the PR of a score of 73 is equal to 79. The interpretation of this value is that a person

holding a score of 73 has a rank of 79 on a 100 point scale. In simple words 79 per cent of the students lie below him in terms of scores.

The above procedure can be diagrammatically shown in Figure 5.1 given below:

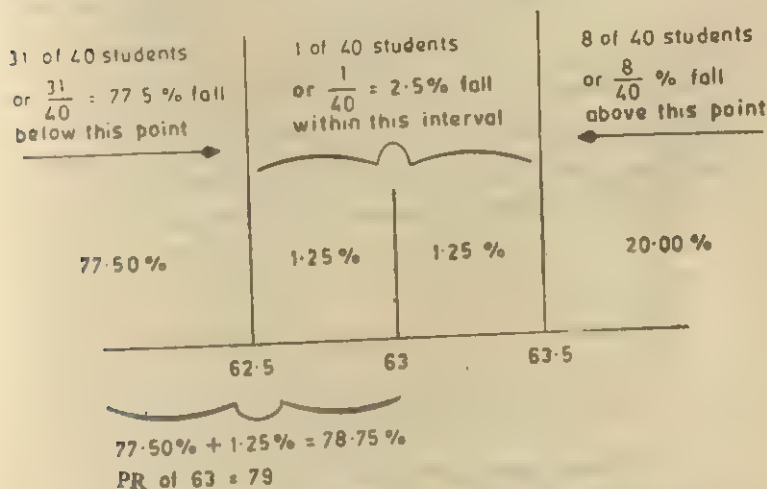


Fig. 5.1. Determination of Percentile Rank Corresponding to the Score Value of 63.

5.4.2 When Duplication Near Percentile Exists

When more than one frequency exists at the given score value, the procedure adopted for the calculation of percentile ranks, in general, is the same as explained above.

For example, we wish to determine PR of score 52. There are three persons holding a similar score of 52 leading to duplication (or triplication) of scores at the percentile. Seventeen ($\frac{17}{40} \times 100$ or 42.5 per cent) students' scores are below the lower limit of 52 which is 51.5. There are three ($\frac{3}{40} \times 100$ or 7.5 per cent) frequencies within this score interval of 51.5 to 52.5. Hence to reach from a score of 51.5 to 52.0, half of this percentage will be added to the percentage covered upto 51.5.

Hence PR of 52 = 42.5 per cent + $\frac{7.5}{2}$ per cent = 46.25 per cent.

By rounding it off, we have 46 as the PR of 52. The above procedure has been presented diagrammatically below in Figure 5.2.

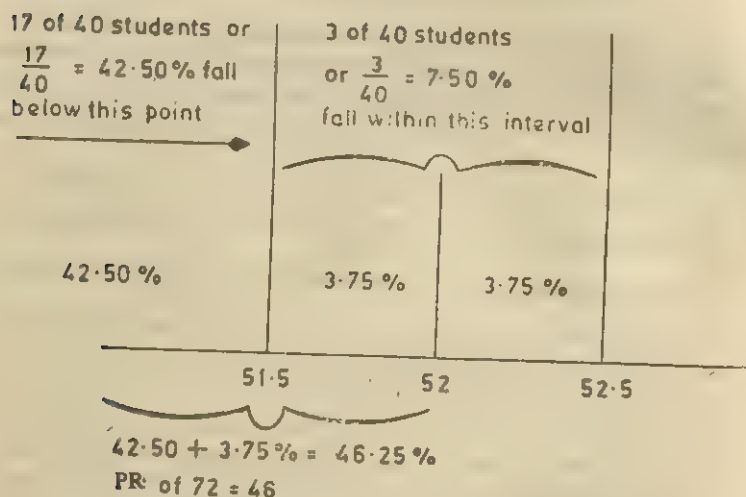


Fig. 5.2. Determination of Percentile Rank Corresponding to a Score value of 52.

5.5 Calculation of Percentiles from the Grouped Data or Frequency Distribution

Calculation of percentile points from a frequency distribution of scores follows the same procedure as discussed in a previous chapter with reference to the calculation of median, Q_1 and Q_3 . The same will be presented here and calculation of several percentile points will be done with reference to the data given below in Table 5.3.

TABLE 5.3

Calculation of P_{10} , P_{20} , P_{30} , P_{40} , P_{50} , P_{60} , P_{70} , P_{80} and P_{90}

Class interval	Exact Limits (X)	Frequency (f)	Commulative frequency (cf)	Percentage
45-49	44.5-49.5	2	50	100.00
40-44	39.5-44.5	3	48	96.00
35-39	34.5-39.5	2	45	90.00
30-34	29.5-34.5	6	43	86.00
25-29	24.5-29.5	8	37	74.00
20-24	19.5-24.5	8	29	58.00
15-19	14.5-19.5	7	21	42.00
10-14	9.5-14.5	5	14	28.0
5-9	4.5-9.5	9	9	18.0

N=50

$$\text{Formula } = P_p = L + \left(\frac{PN - F_b}{f_w} \right) \times i \quad (5.1)$$

where, P_p = the required percentile point P = the proportion of the distribution wanted N = No. of cases L = exact lower limit of the class interval in which P_p falls. F_b = Cumulative frequency below the CI containing P_p f_w = frequency within the CI containing P_p i = size of class interval.For the given Data, $N=50$, $i=5$ Calculation of P_{10} $P=10\%$, $PN=10\%$ of $50=5$, $L=4.5$, $F_b=0$, $f_w=9$

$$P_{10} = 4.5 + \frac{5-0}{9} \times 5 = 4.5 + 2.78 = 7.28$$

2. Calculation of P_{20}

$$P = 20\%, PN = 20\% \text{ of } 50 = 10, L = 9.5, F_b = 9, f_w = 5$$

$$P_{20} = 9.5 + \frac{10 - 9}{5} \times 5 = 9.5 + 1 = 10.5$$

3. Calculation of P_{30}

$$P = 30\%, PN = 30\% \text{ of } 50 = 15, L = 14.5, F_b = 14, f_w = 7$$

$$P_{30} = 14.5 + \left(\frac{15 - 14}{7} \right) \times 5 = 14.5 + \frac{1}{7} \times 5 = 14.5 + .71 = 15.21$$

4. Calculation of P_{40}

$$P = 40\%, PN = 40\% \text{ of } 50 = 20, L = 14.5, F_b = 14, f_w = 7$$

$$P_{40} = 14.5 + \frac{20 - 14}{7} \times 5 = 14.5 + \frac{6}{7} \times 5 = 14.5 + 4.29 = 18.79$$

5. Calculation of P_{50}

$$P = 50\%, PN = 50\% \text{ of } 50 = 25, L = 19.5, F_b = 21, f_w = 8$$

$$P_{50} = 19.5 + \frac{25 - 21}{8} \times 5 = 19.5 + \frac{4}{8} \times 5 = 19.5 + 2.5 = 22$$

6. Calculation of P_{60}

$$P = 60\%, PN = 60\% \text{ of } 50 = 30, L = 24.5, F_b = 29, f_w = 8$$

$$P_{60} = 24.5 + \frac{30 - 29}{8} \times 5 = 24.5 + .62 = 25.12$$

7. Calculation of P_{70}

$$P = 70\%, PN = 70\% \text{ of } 50 = 35, L = 24.5, F_b = 29, f_w = 8$$

$$P_{70} = 24.5 + \frac{35 - 29}{8} \times 5 = 24.5 + 3.75 = 28.25$$

8. Calculation of P_{80}

$$P = 80\%, PN = 80\% \text{ of } 50 = 40, L = 29.5, F_b = 37, f_w = 6$$

$$P_{80} = 29.5 + \frac{40 - 37}{6} \times 5 = 29.5 + 2.5 = 32.0$$

9. Calculation of P_{90}

$$P = 90\%, PN = 90\% \text{ of } 50 = 45, L = 34.5, F_b = 43, f_w = 2$$

$$P_{90} = 34.5 + \frac{45 - 43}{2} \times 5 = 34.5 + 5 = 39.5$$

In Table 5.3, several percentile points have been calculated by using formula (5.1). The details of the calculations have also been presented in the latter half of the same table. However, for further clarification and understanding of the concept of percentile, the calculation of P_{40} is discussed here. Here $PN=20$ (40 per cent of $50=20$). From the frequency distribution Col. (4), it is evident that 14 cases are covered upto a score of 14.5, the exact upper limit of the class interval of 10-14 and also the exact lower limit of the class interval 15-29 in which P_{40} lies. Hence, to reach a score point below which 20 cases would lie, we need 6 cases out of the 7 contained in the score interval of 15-29. The size of the class interval is 5. Now we need to add $(6/7) \times 5 = 4.29$ score units to the lower limit of 14.5 upto which 14 cases had been covered.

$$\text{Hence } P_{40} = 14.50 + 4.29 = 18.79.$$

A similar value has been obtained by using Formula (5.1) as shown in Table 5.3.

P_0 is the exact lower limit of the lowest interval, and P_{100} , the exact upper limit of the top interval. These are called the *limiting points*. In Table 5.3, $P_0=4.5$ and $P_{100}=49.5$. These values can be picked up simply by inspection.

The student, by slight understanding, can also pick up some values of percentile points by inspection only.

For example in Table 5.3, $P_{28}=14.5$ (upper limit of CI 10-14 upto which 14 cases or 28 per cent cases are covered);

$P_{42}=19.5$ (upper limit of CI 15-29 upto which 21 cases or 42 per cent cases are covered)

Similarly, $P_{58} = 24.5$; $P_{74} = 29.5$; $P_{86} = 34.5$; $P_{90} = 39.5$; $P_{98}=44.5$. The student may check the correctness of these values by using Formula (5.1).

5.6 Calculation of percentile Ranks from the Grouped Data

The calculation of percentile ranks requires the reverse of the procedure used for the calculation of percentile points. The cumulative percentages shown in column 5 of Table 5.3 are the percentile ranks corresponding to the exact top limits of the intervals. Thus, 18.0 is the PR corresponding to the percentile

point of 9.5, the upper limit of the class interval 5-9. Similarly, 28.0 is the PR corresponding to the percentile point of 14.5, the upper limit of the class interval. Other values are:

PR of score 19.5=42.0; PR of score 24.5=58.0; PR of score 29.5=74.0; PR of score 34.5=86.0; PR of score 39.5=90.0; etc.

These values of PR's have been picked only by the inspection of Col. (5) of Table 5.3. Other values may require some further calculations.

PR of any score can be obtained by interpolation as illustrated below:

Suppose we wish to obtain PR of a score of 22. The score 22 falls within the interval with exact limits 19.5 and 24.5. Hence, 22 is 2.5 score units above the lower limit of this class interval. The lower limit has a percentile rank of 42.00 and the upper limit, 58.0 with a distance of $58.0 - 42.0 = 16$ units on the percentile rank scale. The number of score units covered is $24.5 - 19.5 = 5$. Hence the number of percentile rank units equal to score units of 2.5 are equal to $(58.0 - 42.0) \frac{2.5}{5} = 8.0$. We now take $42.0 + 8.0 = 50.0$ as the percentile rank of the score 22. Diagram given below will clarify the calculation. Percentile Ranks are reported in whole numbers and hence be rounded to the nearest integer. In this example, the PR has turned out to be an integer leaving no scope for rounding.

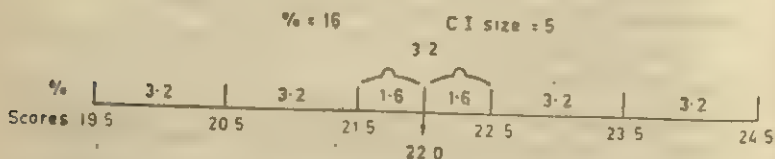


Fig. 5.3. Interpolation for the calculation of PR

Note the steps required in the calculation of PR's.

1. Find the class interval containing the score X whose percentile rank is required.
2. Find the exact lower limit of this class interval.
3. Calculate the difference between X and this lower limit by subtracting the lower limit from X .

4. Divide this difference by the size of the class interval and multiply by the percentage within the interval.
5. Add this to the percentile rank corresponding to the lower limit of the interval (or the percentage covered up to the lower limit of the interval).

5.7 The Cumulative Percentage Curve or Ogive

The cumulative percentage curve or ogive is different from the cumulative frequency graph in one important aspect. In an

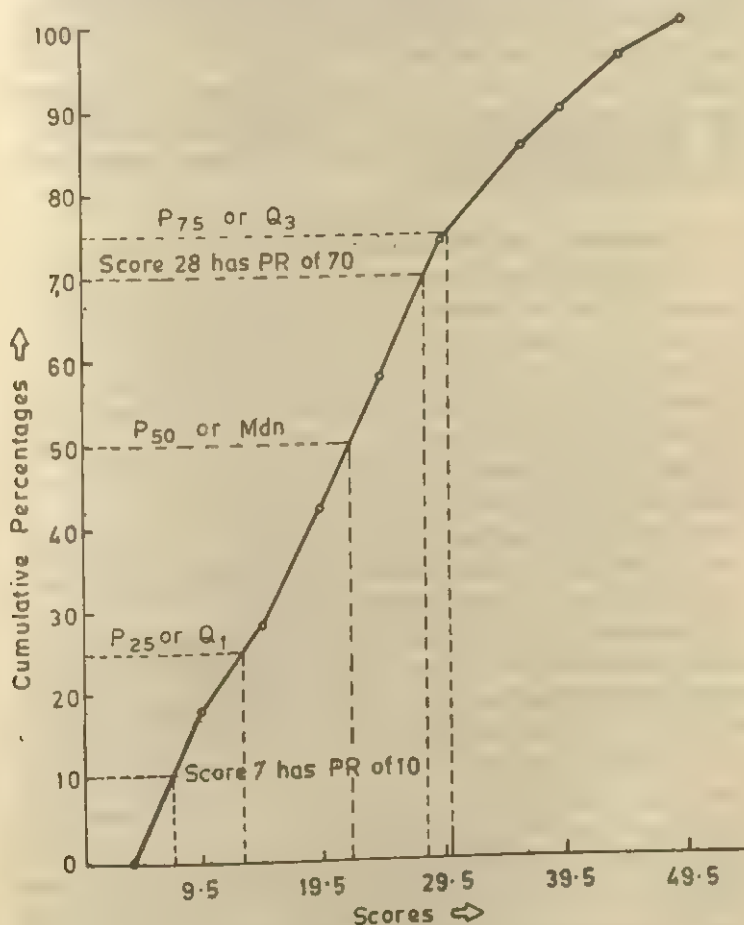


Fig. 5.4. Cumulative Percentage Curve for the Calculation of Percentiles and PR's.

ogive, the frequencies are expressed as cumulative per cents of N on the Y-axis instead of as cumulative frequencies. Table 5.3 shows the process of converting the cumulative frequencies (Col. 4) into cum. percentages (Col. 5). This conversion can be carried out by dividing each cum. frequency by N and multiplying it by 100. The multiplication can be performed simply by shifting the decimal points two spaces to the right.

The curve in Figure 5.4 is an Ogive plotted from the data in column (5) of Table 5.3. Exact interval limits have been laid off on the X-axis and a scale of 10 equal distances, each representing 10 per cent of the distribution has been marked off on the X-axis. The first point on the ogive is placed 18.0 Y-units above 9.5; the second point is 28.0 Y-units just above 14.5 etc. The last point 100 Y units is above 49.5, the exact upper limit of the highest class interval.

5.7.1 Percentiles and Percentile Ranks from Ogive

Percentiles and PR's may be determined quickly and fairly accurately from an ogive. In Figure 5.4, an ogive based on the data of Table 5.3 has been shown. The median or P_{50} , Q_1 or P_{25} , Q_3 or P_{75} and a few other percentiles have been marked off on the ogive. To obtain median or P_{50} , draw a line from 50 on the Y scale parallel to the X axis and from where this line cuts the curve, drop a perpendicular on the X axis. Read this point of intersection on the X axis. This value is 21.83 which is the median. Similarly locate P_{25} and P_{75} on the Y scale and determine X values by drawing perpendiculars. These values are 12.83 and 30.00 respectively.

Determination of percentile ranks requires a reversal of the above procedure. Here, we start off with a score on X axis and draw a perpendicular on it. From the point where it meets the curve, we draw a line parallel to the X axis. The point where it cuts the Y axis is read as the required percentile rank. Percentiles and percentile ranks read from an ogive will often be slightly in error yet accurate enough to serve the practical purposes for which these are generally used. However, when the diagram is fairly large, the scale divisions precisely marked and the curve is carefully drawn, percentiles and PR's can be read more accurately.

The ogive has several other uses. It can be used to compare two or more groups on some variables of interest to the researcher. For this purpose, the scores of both the groups are plotted upon the same coordinate axes. Differences and similarities between the two distributions at all the points of the scale can then be studied.

The percentile points also serve as percentile norms for the comparison of a person's score with reference to his group. These are particularly useful in dealing with educational achievement examinations. Intra-student comparisons on more than one subject are also possible.

5.8 Standard Scores

A standard score is a deviation from the mean divided by the standard deviation. It is denoted by z and the formula is

$$z = \frac{X - M}{\sigma} \quad (5.2)$$

in which, z = standard score

X = raw score

M = mean of raw scores

σ = standard deviation of raw scores.

Standard scores have a mean equal to zero; and SD, equal to unity or one. Thus, the mean serves the purpose of the origin and SD, the unit of measurement. Thus, a particular score value is z standard deviation units above or below the mean.

Transformation of raw scores into z scores do not, in any way, change the other characteristics of the distribution like, skewness and kurtosis. The shape of the distribution remains absolutely unchanged. It does not, in any way, change the proportionality of the scale intervals. It means that the relative distances between the score values remain unchanged under a standard score transformation. The procedure of calculation of z scores is shown in Table 5.3A on next page.

Thus each pupil's level of excellence is expressed as so many standard deviation units above or below the mean of the comparison group. Standard scores have essentially the same meanings from one test to another test.

TABLE 5.3A
Computation of Standard Scores

Test	Person	Raw Score	Mean	Standard deviation (σ)	$M - X$	$\frac{M - X}{\sigma}$ or z
Number Ability	Ram	62	56	8	62 - 56 = 6	6/8 = .67
	Sham	42	56	8	42 - 56 = -14	-14/8 = -1.67
Spellings	Ram	68	70	6	68 - 70 = -2	-2/6 = -.33
	Sham	58	70	6	58 - 70 = -12	-12/6 = -2.0

Interpretation: Ram is .67 SD units above the mean of the group on Numerical Ability Test while Sham is 1.67 SD units below the group. On spellings Test, both the persons are below the average by .33 and 2.0 standard deviation units.

Inter-test comparisons reveal that Ram, though having a higher raw score on Spellings Test yet he is .33 standard deviation units below the mean while with a smaller raw score, he is .67 units above the group mean on Numerical Ability Test.

Standard scores in standard deviation units are quite satisfactory for several purposes but they involve two difficulties:

- (i) Plus and minus signs are used which can be miscopied, overlooked, or misunderstood, and
- (ii) Decimal points are involved which may be misplaced.

Hence, it has been suggested that a mean of 50 instead of zero and an SD of 10 instead of one, be used to avoid both the negative signs and the decimal points. For example, the raw scores given in Table 5.3 can be changed into standard scores with $M=50$ and $\sigma=10$, through the procedure shown in Table 5.4.

TABLE 5.4

Computation of Standard Scores with $M=50$, $SD=10$

Formula

$$z = 10 \left(\frac{X - M}{\sigma} \right) + 50 \quad (5.3)$$

Numerical Ability Ram $z = 10 \left(\frac{62 - 56}{8} \right) + 50 = 56.7$ or 57

Sham $z = 10 \left(\frac{42 - 56}{8} \right) + 50 = 32.5$ or 33

Spellings Ram $z = 10 \left(\frac{68 - 70}{6} \right) + 50 = 46.67$ or 47

Sham $z = 10 \left(\frac{58 - 70}{6} \right) + 50 = 30$

The values of mean and standard deviation can be set up arbitrarily for the purpose of conversion. However, we may come across $M=50$ and $SD=15$; and $M=500$ and $SD=100$ more often in the literature than any other values.

5.9 The Stanine Scale

The stanine (abbreviated from standard nine) is a condensed or coarser form of the normalized T scale. Only nine score categories or groups are formed and the integer values 1 to 9 are

assigned to each. The base line of the normal curve is divided into 9 equal divisions in terms of standard deviation units and thus the following percentages of area is obtained.

TABLE 5.5
The Stanine Score System

<i>Stanine scale</i>	<i>in each interval (rounded)</i>	<i>Cum. %'s</i>
9	4	100
8	7	96
7	12	89
6	17	77
5	20	60
4	17	40
3	12	23
2	7	11
1	4	4

If a set of scores is ordered from the lowest to the highest, the lowest 4 per cent assigned a score of 1, the next lowest 7 per cent a score of 2, and the process continued till the top 4 per cent get a score of 9, as shown in table 5.5. The transformed scores are roughly normal and form a stanine scale. The stanine scale has mean = 5, SD = 1.96.

A stanine of 5 covers the interval -2.5 to $+2.5$ in standard deviation units. The relationship of stanines with the σ scores and area per cent is given in Figure 5.5.

A stanine scale provides a quick method of converting scores to an approximate normal form. The grouping, although coarse, is sufficiently refined for many practical purposes.

5.10. The T-Scale

Normalized standard scores are generally called T scores. McCull (1939) devised T scale for the first time which became

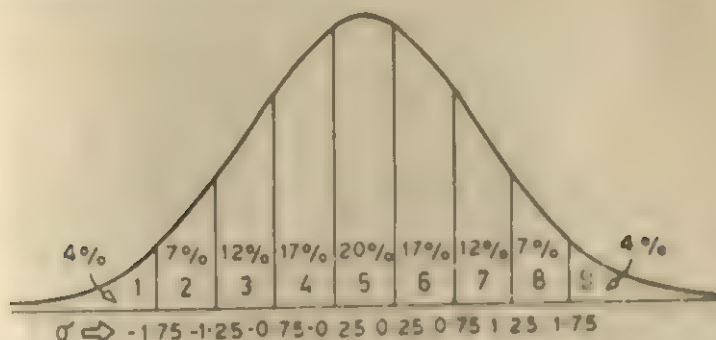


Fig. 5.5. Normal scale showing standard deviation intervals and percents in each score from 1 to 9.

very popular later on. In the standard scores or z scores, the mean is at zero and $\sigma = 1.00$. The point of reference is then zero and the unit of measurement is 1. However, in T scores, a mean of 50 and σ of 10 are used. See Figure 5.6 for a comparison of various scales.

Only slight changes are needed to convert the z scale into a T scale. The T scale begins at -5σ and ends at $+5\sigma$. But σ is multiplied by 10 so that the mean is 50 and the other divisions are 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. The T scale thus ranges from 0 to 100, and its unit is 1, and mean, 50. Ability ranges beyond -3.5σ to $+3.5\sigma$ are rare to find hence in actual practice T scales range from about 15 to 85. Calculation of T scores has been illustrated on page 99 in Table 5.6.

Steps in the calculation of T scores

- (1) List up the class intervals in Col. (1), midpoint of each CI in Col. (2); and f in Col. (4).
- (2) Calculate cum. frequencies and enter in Col. (4).
- (3) Col. (5) shows the cum. f to the midpoints. These are frequencies below a particular CI plus f for that CI. For example, CI 5-9 has 2 frequencies in it and zero frequencies below it. Hence cum. f to midpoint for this CI = $0 + 2 + 1.2 = 1.2$. Similarly, for CI 10-14,

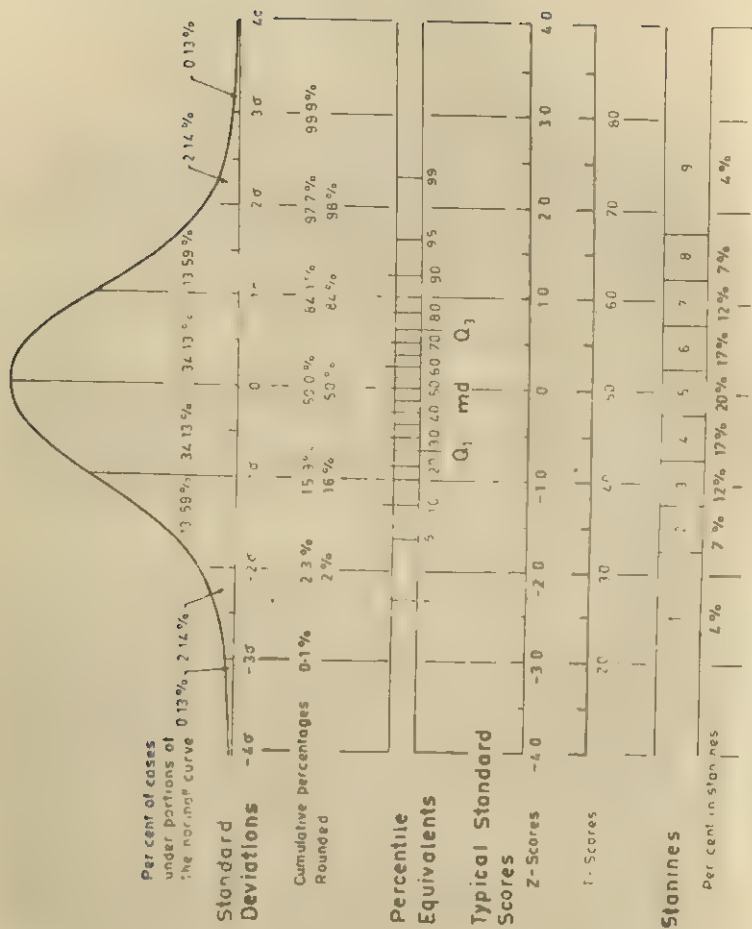


Fig. 5.6. Various types of standard score scales in relation to percentiles and the normal curve.

TABLE 5.6
Computation of T Scores

Test-Scores	Midpoint	f	Cum. f	Cum. f below score + $\frac{1}{2}$ on given score (to midpoint)	Cum. % to midpoint	Normal standard deviation unit z	$\frac{T \text{ Score}}{z \times 10}$	$\frac{T \text{ Score}}{z \times 10} + 50$	T Scores from Appendix
45-49	47	1	50	49.5	99.0	2.33	23.3	73.3	73
40-44	42	4	49	47	94.0	1.56	15.6	65.6	66
35-39	37	7	45	41.5	83.0	.95	9.5	59.5	60
30-34	32	8	38	34	68.0	.47	4.7	54.7	55
25-29	27	10	30	25	50.0	0.00	0.0	50.0	50
20-24	22	8	20	16	32.0	-0.47	-4.7	45.3	45
15-19	17	6	12	9	18.0	-0.92	-9.2	40.8	41
10-14	12	4	6	4	8.0	-1.40	-14.0	36.0	36
5-9	7	2	2	1.0	2.0	-2.05	-20.5	29.5	30

two frequencies below and 4 in it, leading to cum. f. upto midpoint equal to $2+4 \times (1/2)=4$, and so on.

- (4) Calculate cum. percentage to midpoint (col. 6) for each CI. These virtually are percentile ranks corresponding to each midpoint.
- (5) Col. (7) carries the values of normal deviate, z , corresponding to each cum. percentage.
- (6) z values have been multiplied by 10, and a constant of 50 added to each to arrive at the value of T scores (Col. 8). For example for CI 10-14, $T=(-1.40 \times 10)+50=36$; for CI 40-44, $T=(1.56 \times 10)+50=65.6$ and so on.
- (7) Col. (9) shows T scores corresponding to each cum. per cent to midpoint, read from Appendix, Table L. It would eliminate the need of calculating Z values and further computational work. Values read from the Appendix are very close to those worked out by actual calculation. However, for a student, it is better to do all the calculations to understand the concept of T scale.

T scale scores have general applicability, a very convenient unit and cover a wide range of ability. The minus and fractional values are eliminated. They are comparable from test to test and provide the same meaning. However, if suspicion about the normality of the trait in the population arises, T scores should not be used. However, it has been seen that when large samples are tested on mental abilities, normality is a reasonable assumption.

Exercises for Practice

- 5.1 What are the limitations of percentiles as measures of relative position and how do standard scores overcome these limitations?
- 5.2 What is the difference between percentile points and percentile ranks?

- 5.3 What are Age Norms and Grade Norms? What are their relative merits and demerits?
- 5.4 Why are "stanines" called so? What purposes do they serve?
- 5.5 Develop a T-score transformation for the following data:

<i>Scores</i>	<i>f</i>
50—54	1
45—49	2
40—44	3
35—39	6
30—34	8
25—29	17
20—24	26
15—19	11
10—14	2
5—9	0
76	

- 5.6 Develop a stanine transformation on the data given in 5.5 above.
- 5.7 Pick up P_{25} , P_{50} , P_{75} , P_{10} , P_{15} , P_{85} and P_{94} from the data given below (No calculations)

<i>Scores</i>	<i>f</i>	<i>cf</i>
80—89	12	200
70—79	18	188
60—69	20	170
50—59	50	150
40—49	50	100
30—39	20	50
20—29	10	30
10—19	14	20
0—9	6	6

N=200

- 5.8 Check the correctness of your results of Q. 5.7 above through calculations.
- 5.9 Compare the relative performance of the three students on Mathematics and History, by using z scores:

<i>Students</i>	<i>Marks</i>	
	<i>History</i>	<i>Maths.</i>
I	52	80
II	63	85
III	42	75
Mean	50	82
SD	8	10

CHAPTER 6

PROBABILITY, BINOMIAL DISTRIBUTION AND NORMAL DISTRIBUTION

Probability theory had its origins in games of chance. Now it has become a fundamental tool of scientific thinking. In general, the interpretation of the data of experiments is in probabilistic terms. The probability theory incorporating different probability models helps the scientist to interpret the relationship between the deductive consequences of theory and the observed data. Several theoretical models, binomial, normal, poisson, hypergeometric etc. are in vogue. However, the first two are more popularly used in educational research. It is so because of their suitability to the data based on educational phenomena.

A definition of probability can follow three approaches—*The subjective or personalistic approach* which is based on statements like, “The probability is high that it will probably rain today”. The second approach, *the formal mathematical approach*, defines the probability of an event as the ratio of the number of favourable cases to the total number of equally likely cases. This usage is based on games of chance, involving cards, dice and coins. The probability of getting a 3 in one throw of dice is $\frac{1}{6}$. In this way, this usage is based on a concept of equally likely cases. The postulate of “equally likely cases” is a theoretical one and is not based on empirical considerations. The third, the *empirical relative frequency approach* considers relative frequencies as the basis of prediction. If a series of N trials is made, and a given event occurs r times, then r/N is the relative frequency. The relative frequency in a sample of observations is an estimate of that parameter.

The three approaches to probability are not incompatible. All three, in fact, often co-exist. While the purely intuitive probability may be an interesting topic of psychological inquiry, the other two approaches are widely used in statistical work, the relative frequency approach being the more consistent of the formal mathematical one.

6.1 Some Fundamental Notions

The concepts which are fundamental to the understanding of probability are described below.

6.1.1 Possible outcomes

In tossing a coin, the number of possible outcomes are two, either head or a tail, i.e. H or T. In tossing two coins, the four possible outcomes are:

First Coin	Second Coin	Description	Symbols
Head	Head	Both coins heads	HH
Head	Tail	First coin head, Second tail	HT
Tail	Head	First coin tail, Second head	TH
Tail	Tail	Both coins tails	TT

Similarly, in tossing three coins, the possible outcomes are HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT.

When a die is thrown, the possible outcomes are six

1, 2, 3, 4, 5, 6.

When two dice are thrown, the number of possible outcomes are 36 and can be listed as shown in Table 6.1.

In drawing a single card from a deck of 52 cards the number of possible outcomes are 52. But in drawing one card from one deck of 52 cards and another card from a different deck, the possible outcomes are $52 \times 52 = 2704$.

With most coins, dice and cards the basic assumption of equal likelihood of all possible outcomes is justifiable. The

TABLE 6.1

List of 36 possible outcomes when two dice are thrown

*I	II	I	II	I	II	I	II	I	II	I	II
1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6	5
1	6	2	6	3	6	4	6	5	6	6	6

*I and II stand for first and second dice

truth of the statement can be viewed by experiment. The probability of an event is denoted by $p(\text{event})$. For instance, the probability of one head in the next coin casting will be $p(H)$, and of one tail $p(T)$.

Probability can be viewed in the form of the number of favourable cases to the total number of equally likely cases. The following example may be viewed with profit.

The total probability for all the possible events as shown above is always unity or 1.00 and can be arrived by adding up the probabilities of the individual events. We must remember that the probability of an event occurring implies that it will not occur (say, equally 1.00 if we think of probability as there). In the above example, we agree with the above probabilities associated with the event pointed on the basis of the equally likely events. However, sometimes the determination of the probability requires empirical approach. For example, to know the probabilities of a person going to the hospital in the population of Delhi to be over the age of 40 would require the use of census data.

TABLE 6.2

Calculation of Probability in different situations

<i>Event</i>	<i>No. of favourable cases</i>	<i>Total No. equally likely</i>	<i>Probability Col. (2)/Col. (3)</i>
<i>One Coin Example</i>			
One Head	1	2	$1/2 = .5$
One Tail	1	2	$1/2 = .5$
			Total = 1.0
<i>Two Coin Example</i>			
Two Heads	1	4	$1/4 = .25$
Two Tails	1	4	$1/4 = .25$
One Head, one tail	2	4	$2/4 = .50$
			Total = 1.00
<i>Three Coin Example</i>			
Three heads	1	8	$1/8 = .125$
Three tails	1	8	$1/8 = .125$
Two heads, one tail	3	8	$3/8 = .375$
One head, two tails	3	8	$3/8 = .375$
			Total = 1.00
<i>One Dice Example</i>			
Number 1	1	6	$1/6 = .167$
Number 2	1	6	$1/6 = .167$
Number 3	1	6	$1/6 = .167$
etc.			

6.1.2 Addition and multiplication rules

The addition theorem states that *the probability that any one of a number of mutually exclusive events will occur is the sum of*

the probabilities of the separate events. Events that cannot happen at the same time are mutually exclusive. In a one coin toss problem, the occurrence of a head is mutually exclusive with the occurrence of a tail as both cannot happen at a time. In a throw of a dice, the probability of obtaining each of a 1, 2, 3, 4, 5 or 6 is $1/6$; what is the probability of obtaining either a 1 or 2 or 4 in a single throw? This can be obtained by adding up the probabilities associated with 1, 2 and 4, i.e. $1/6 + 1/6 + 1/6 = 3/6$ or $1/2$. In tossing two coins, four possible events, HH, HT, TH and TT are possible. What is the probability of obtaining either two heads or two tails? The probability of each of these events is $1/4$. Hence the probability of obtaining either two heads or two tails is $1/4 + 1/4 = 1/2$.

The multiplication theorem states that the probability of the joint occurrence of two or more independent events is the product of their separate probabilities. When a single coin is tossed twice, the probability of getting a head on the first is $1/2$, and on the second toss, $1/2$. The probability of getting two heads is therefore $1/2 \times 1/2 = 1/4$. In the same manner, we could determine that the probability of getting three heads from tossing a single coin three times would be $1/2 \times 1/2 \times 1/2 = 1/8$. What is the probability of obtaining 6's in rolling two dice? The probability that the first die is a 6 is $1/6$. The probability that the second die is a 6 is also $1/6$. Hence the probability that both dice are 6's is $1/6 \times 1/6 = 1/36$.

6.1.3 Permutations and combinations

Sometimes questions of the following kind are asked: 'In how many ways can five books be arranged on a shelf?' 'In how many ways can six persons be seated at a table?' The answer lies in calculating the number of possible arrangements. Any arrangement is called a *Permutation*. Order is the essential idea here and a different order is a different permutation. With two objects A and B, two arrangements, or permutations, are possible, AB and BA. Three objects A, B and C can provide six arrangements or permutations: ABC, ACB, BAC, BCA, CAB, and CBA. In general, if there are n distinguishable objects, the number of permutations of these objects taken n at a time are given by $n!$ read as n factorial. Factorial of any

numt is the product of all integers from that number to 1 e.g. $4! = 4 \times 3 \times 2 \times 1 = 24$. The value of n in the three objects example above is 3. Hence the number of permutations equals $3!$ or $3 \times 2 \times 1 = 6$. The number of possible arrangements or permutations of 5 books equals $5!$ or $5 \times 4 \times 3 \times 2 \times 1 = 120$; and of six persons, $6!$ or $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

If arrangements require r to be taken at a time, when r is less than n , the formula for the calculation of permutations is

$${}_nP_r = n(n-1) \dots (n-r+1) = \frac{n!}{(n-r)!}$$

For example, the number of arrangements of ten objects taken three at a time is

$${}_{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 720$$

Combinations are the arrangements of objects when the order in which they are arranged is ignored.

Given the objects A, B, C and D, the number of permutations of two from this set is $4!/(4-2)! = 12$. These are—AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD and DC. It is evident that each arrangement occurs in two different orders. If the order of arrangement of each pair of objects is ignored, we have the number of combinations. Obviously, the number of combinations, in this example, would be reduced. In general, the formula for calculating the number of combinations is:

$${}_nC_r = \frac{n!}{r!(n-r)!} \quad (6.1)$$

read as: n factorial upon r factorial into $(n-r)$ factorial. ${}_nC_r$ stands for the number of combinations of n things taken r at a time; other terms are factorials. The number of combinations of 10 things taken 2 at a time is $10!/2!(10-2)! = 45$. The number of combinations of n things taken n at a time is obviously 1, because there is only one way of picking all n objects if the order of their arrangement is ignored.

6.2 The Binomial Distribution

Suppose we have the hypothesis that a student taking a true-false test will respond to each item by tossing a coin.

If we assume that 50 per cent of the times the toss will result in correct answers and the rest of the 50 per cent times, in incorrect answers, we may say that the probability of making a correct answer is $1/2$ and is equal to the probability to making an incorrect answer which is also $1/2$. Suppose further, that the test contains 10 true-false items. The questions that may be asked are: what is the probability of the student obtaining all the 10 items correct; or all the 10 items incorrect; or 7 answers correct and 3 answers incorrect. In such situations, binomial distribution provides the answer. To answer the last question, we may use the formula:

$${}_nC_r p^r q^{n-r} = \left(\frac{n!}{(n-r)!r!} \right) (p)^r (q)^{n-r} \quad (6.2)$$

where, ${}_nC_r$ is the number of combinations of n things taken r at a time:

p is the probability of getting a correct answer

q the probability of getting an incorrect answer

n is the total number of questions

r is the number of correct answers desired.

Substituting the numerical values in the formula, we obtain

$$\begin{aligned} {}_{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^{10-7} &= \frac{10!}{(10-7)!(7)!} \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 \\ &= \frac{120}{1024} = .117 \end{aligned}$$

Similarly, we could use the above formula to obtain the probability of the student getting any particular score, ranging from 10 to 0 correct answers.*

The binomial for n things can be expanded as follows:

$$\begin{aligned} (p+q)^n &= p^n + np^{n-1}q + \frac{n(n-1)}{(1)(2)} p^{n-2}q^2 + \dots \\ &\quad + \frac{n(n-1)(n-2)}{(1)(2)(3)} p^{n-3}q^3 + \dots + q^n \end{aligned} \quad (6.3)$$

*It is customary to consider $0! = 1$.

For the problem of 10 true-false items mentioned above, the binomial expansion will be as given below:

$$(p+q)_{10} = p^{10} + 10 p^9 q + 45 p^8 q^2 + 120 p^7 q^3 + 210 p^6 q^4 + 252 p^5 q^5 + 210 p^4 q^6 + 120 p^3 q^7 + 45 p^2 q^8 + 10 p q^9 + q^{10}$$

The value of p and q which is $1/2$ in each case can also be inserted. The fourth term then would be $(120)(1/2)^7(1/2)^3$

The exponent of p in each of the terms of the binomial expansion as in Formula (6.3) indicates the number of items correct (successes) and that of q indicates, the number of items incorrect (failures). The numerical coefficients represent the number of ways in which each of the combinations of successes and failures may occur.

The rules for expanding the binomial $(p+q)^n$ are summarized below:

- (1) Each term in the binomial consists of the product of a numerical coefficient and a power of p and power of q .
- (2) The first term always has a numerical coefficient of 1 which is understood and hence not written; the power of p in the first term is always n , and the power of q is zero; since $q^0 = 1$, q does not appear. Thus the first term always is p^n .
- (3) In each succeeding term, the power of p decreases by 1 in regular order, while the power of q increases by 1 in regular order until the final term, q^n , is obtained.
- (4) The product of the numerical coefficient and the power of p in any given term, divided by 1 plus the power of q in that term, will give the numerical coefficient of the term that follows.

For example, the numerical coefficient 45, of the third term, has been obtained by multiplying the coefficient of the second term by its power of p and then dividing by one plus the power of q . Thus

$$\frac{(10)(9)}{1+1} = \frac{90}{2} = 45.$$

The numerical coefficient for any combination of correct and incorrect answers can be obtained by the formula

$${}_nC_r = \frac{n!}{(n-r)! (r)!} \quad (6.4)$$

In the above example, with $n=10$ items; and no. of correct answers, $r=3$, the numerical coefficient will be

$${}_{10}C_3 = \frac{10!}{(10-3)! (3)!} = 45$$

The coefficients for n upto 10 are given in Table 6.4. It may be noted that any entry in a given row consists of the sum of the coefficients to the right and left of the entry in the row directly above. Thus, the entries for $n=11$ can be obtained from the entries for $n=10$. They would be 1, 11, 55, 165, 330, 462, 462, 330, 165, 55, 11 and 1. Since the binomial is symmetric, the values of the numerical coefficients to the left and to the right of the middle term/terms are equal.

TABLE 6.4

**The Binomial Coefficients of $(p+q)^n$
Pascal's Triangle**

n	<i>Binomial coefficients</i>										<i>sum</i>
1				1		1					2
2			1		2		1				4
3			1		3		3		1		8
4			1		4		6		4		16
5			1		5		10		10		32
6			1		6		15		20		64
7			1		7		21		35		128
8			1		8		28		56		256
9			1		9		36		84		512
10			1		10		45		120		1024

If we tested N students with our true-false test and, if we still assume that each student answered each item by flipping a coin, that is, by chance, then we may readily determine the

number of students expected to obtain each possible score
Formula (6.3) would thus become:

$$N(p+q)^n = Np^n + N(np^{n-1}q) + N\left(\frac{n(n-1)}{(1)(2)}p^{n-2}q^2\right) + \dots + Nq^n \quad (6.5)$$

In which: N = The number of students tested; n = the number of items in the test; p = probability of a correct response to a single item; $q = 1 - p$.

Since the sum of the numerical coefficients for $n=10$, as given in the Pascal's triangle is 1024, we can, for simplicity, take $N = 1024$. The probabilities and N 's for various events are given in Table 6.5.

TABLE 6.5

**The Binomial Distribution $(p+q)^n$ and $N(p+q)^n$
with $p = .5$, $n = 10$ and $N = 1024$**

Score Number correct	Score proportion correct	Probability	Expected number of students, f
10	1.0	$1/1024 = .001$	1
9	.9	$10/1024 = .010$	10
8	.8	$45/1024 = .044$	45
7	.7	$120/1024 = .117$	120
6	.6	$210/1024 = .205$	210
5	.5	$252/1024 = .246$	252
4	.4	$210/1024 = .205$	210
3	.3	$120/1024 = .117$	120
2	.2	$45/1024 = .044$	45
1	.1	$10/1024 = .010$	10
0	.0	$1/1024 = .001$	1
Σ		1.000	1024

The mean of the binomial distribution is given by the formula

$$m=np \quad (6.6)$$

where, m = the mean number of correct responses of the binomial $(p+q)^n$

n = the exponent of $(p+q)$

p = the probability of having an item correct.

In our case, where $n=10$,

$$m=(10)(.5)=5.0$$

The variance and standard deviation of the binomial are given by the formulas

$$\sigma^2=npq \quad (6.7)$$

$$\sigma=\sqrt{npq} \quad (6.8)$$

where σ^2 = the variance of the binomial distribution $(p+q)^n$

n = the exponent of $(p+q)$

p = the probability of having an item correct

$q=1-p$ or the probability of having an item incorrect.

Substituting our values in the above formulas, we have

$$\sigma^2=(10)(.5)(.5)=2.5$$

$$\sigma=\sqrt{(10)(.5)(.5)}=\sqrt{2.5}=1.58$$

The above formulas are in terms of n or frequency of correct responses or in terms of the scores of col. (1) of Table 6.5. The m and σ of the binomial in terms of proportion of correct responses (col. 2 of Table 6.5) will be given by the formulas:

$$m=p \quad (6.9)$$

$$\sigma=\sqrt{\frac{pq}{n}} \quad (6.10)$$

6.3. The Normal Distribution

The normal distribution is a special case of the binomial distribution with $p=.5$, and a sufficiently large N . As N grows infinitely large, the normal and binomial probabilities become

identical for any interval. It also depicts a tendency that test scores always tend to be distributed around the averages. On a test with items of average difficulty level, many students of an unselected group, will obtain average or near average scores. The number of students in the two tails of the distribution go on decreasing, as we move away from the mean.

The normal distribution, mathematically, is an approximation of the type of distribution generated by tossing coins. If 10 coins are tossed simultaneously a large number of times, the most frequently occurring combination of results would be 5 heads and 5 tails. Other combinations would be fewer and as one proceeds towards 10 heads and no tails, or 10 tails and no heads, the frequencies will go on decreasing. The occurrence of the extreme combinations (10 heads, and 10 tails 0 heads) would be very rare. If the frequencies with which each combination appears are plotted on a graph, an approximately normal distribution will be obtained. Normal distribution is a mathematically theoretical concept or model yet it fits into several real situations for explanation. The characteristics or salient properties of the curve are given below:

DeMoivre (1733) first developed the equation of the curve. This concept was further developed and perfected by Gauss and Leplace.

6.3.1. Properties of the Normal Curve

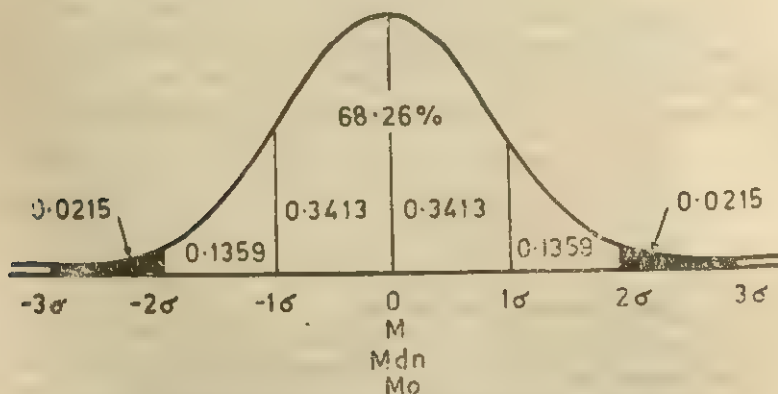


Fig. 6.1. Different proportions of Area under the Normal curve.

1. It is a bell-shaped curve because of its characteristic roundness at the top and inflections on each side.
2. The tails of the curve are asymptotic to the base line. It means that the tails of the curve theoretically approach the baseline but never touch it.
3. The mean, mode and median, in a normal curve coincide or fall at the same point. They have the same numerical values. It means in a normal curve, Mean = Median = Mode.
4. It is symmetrical about the mean and hence the area below the mean is equal to the area above the mean. Thus, it is bilateral.
5. For practical purposes, the baseline of the curve is divided into six sigma distances from -3σ to $+3\sigma$. Most of the cases (i.e. 99.73%) are covered, within $\pm 3\sigma$ from the mean. Very few cases deviate by more than 3σ above or below the mean.
6. The curve is unimodal.
7. The maximum ordinate of the curve occurs at the mean or where $z=0$. The highest of the ordinate at this point is 0.3989. The heights of the ordinates at 1σ , 2σ and 3σ are .2420, .0540 and .0044 respectively.
8. The area under the curve represents the total frequency (N) of the distribution.
9. The points of inflection of the curve occur at points plus and minus one σ unit above and below the mean. It means the curve changes from convex to concave in relation to horizontal axis at these points.
10. In a normal curve.

Quartile deviation, $Q = \text{Probable Error} = 0.6745 \sigma$

Mean deviation, $AD = 0.7979 \sigma$

Skewness = 0

Kurtosis = .263

11. The normal curve is defined by the equation :

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{(-x^2)/2\sigma^2} \quad (6.11)$$

(Symbols explained in the next section)

6.3.2 The Equation for the Normal-Distribution Curve

The normal distribution curve is described by the following general mathematical equation :

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{(-x^2)/2\sigma^2} \quad (6.12)$$

where, Y = frequency for any given point of the baseline

N = number of observations

σ = SD of the distribution

π = 3.1416 (Approx.); A mathematical constant

e = 2.718 (Approx.); A mathematical constant

x = deviation of a score from the mean; $(X - \bar{M})$

The equation leads to the following features of the curve:

1. With N and σ fixed, all elements, except x , are constants.
2. Since x is squared, the negative or the positive values of x will yield the same value of Y , hence leading to the symmetry of the curve.
3. Since x^2 has a negative sign, as x increases, the exponent of e decreases and Y also decreases.
4. If $x = 0$, exponent of e becomes zero, and the value of e with the exponent 0 reduces to 1. The equation then reduces to $Y = \frac{N}{\sigma\sqrt{2\pi}}$. The value of Y is at a maximum at this point.

6.3.3 The Unit Normal Curve

Tables of areas of normal curve have been prepared based on $N=1$, and $\sigma=1.0$. The equation (6.12) in these circumstances reduces to

$$Y = \frac{1}{\sqrt{2\pi}} e^{(-z^2/2)}$$

The total area is considered to be unity or one and all frequencies are proportions. The mean of distribution remains at zero.

6.3.4 Areas Under the Normal Curve

The normal-curve tables (Table A) are generally limited to the areas under the Unit Normal Curve, with $N = 1$, $\sigma = 1$. In case when the values of N and σ are different from these, the measurements or scores should be converted into sigma scores or standard scores or Z-scores. The process is as follows:

$$z = \frac{x}{\sigma_x} = \frac{X - M_x}{\sigma_x} \quad (6.13)$$

In which, z = standard score

x = deviation of the raw score from the mean

M_x = Mean of X scores

σ_x = SD of X scores

The tables of areas of Normal curve are then consulted to find out the proportion of area between mean of the curve and the z . While consulting Normal Curve Area tables, the following points should always be kept in mind to avoid error:

1. Everything i.e. scores or observations must be converted into standard measures i.e. z scores as shown above.
2. The mean of the curve is always the reference point, and all the values of areas are given in terms of distances from the mean which is zero.
3. The area in terms of proportion can be converted into percentage by multiplying it by 100 or by simply shifting the decimal two places to the right.
4. While consulting tables, absolute value of z (ignoring sigma) should be taken. However, a negative value of z shows that the score and the area lie below the mean and this fact should be kept in mind while doing further calculations on the area. A positive value of z shows that the score and hence the area also lies above the mean. Fig. 6.1 shows the distribution of areas under the normal curve within the limits of some selected σ units on the baseline.

The table given below depicts the proportions of area between mean and various z scores and the learner should study it carefully and see the various relationships, that exist in these

values. The areas for the positive and negative values of a z score are the same. Proportions of areas have been converted into percentages simply by shifting the decimal points two places to the right. While the baseline distance from one z value to the next is equal, the relative differences in the areas are not equal.

TABLE 6.6

Normal Probability Curve Area Values for Given z Values

z values	<i>Area from Mean</i>		z values	<i>Area from Mean</i>	
	<i>Proportion</i>	<i>Percentage</i>		<i>Proportion</i>	<i>Percentage</i>
+0.5	.1915	19.15	-0.5	.1915	19.15
+1.0	.3413	34.13	-1.0	.3413	34.13
+1.5	.4332	43.32	-1.5	.4332	43.32
+2.0	.4772	47.72	-2.0	.4772	47.72
+2.5	.4938	49.38	-2.5	.4938	49.38
+3.0	.4987	49.87	-3.0	.4987	49.87

*Normal Curve Area tables show only positive values of z .

Study Table 6.7 also carefully and visualize the various relationships in terms of z values and areas.

TABLE 6.7.

Area under Normal Probability Curve between Given Limits

<i>Limits (in z-scores)</i>	<i>Area (in percentage)</i>
Above +1.0	$(50 - 34.13) = 15.87$
Below +1.0	$(50 + 34.13) = 84.13$
Between +0.5 and +0.75	$(27.34 - 19.15) = 8.19$
Between -0.5 and +0.5	$(19.15 + 19.15) = 38.30$
Above +0.75	$(50.00 - 27.34) = 22.66$
Below +2.00	$(50.00 + 47.72) = 97.72$

Above +2.00	$(50.00 - 47.72) = 2.28$
Below +2.5	$(50.00 + 49.38) = 99.38$
Above +2.5	$(50.00 - 49.38) = 0.62$
Between -2.0 and +1.5	$(47.72 + 43.32) = 91.04$
Between +0.4 and -0.2	$(15.54 + 7.93) = 23.47$
Between +0.8 and -0.6	$(28.81 - 22.57) = 6.24$
Between -1.65 and -1.25	$(45.05 - 39.44) = 5.61$

6.3.5 Problems and Numericals on Normal Distribution

Normal distribution has been found very useful in problems involving inferences. Several types of problems can be solved by using normal curve as a theoretical model. However, the following five types of problems will be illustrated through numerical solutions:

- (1) Determination of the percentage of cases within given limits of scores.
- (2) Determination of the limits of scores which include a given percentage of cases [converse of type (1) above].
- (3) Comparison of two distributions in terms of overlapping.
- (4) Determination of the relative difficulty of test questions, problems and other test items.
- (5) Division of a given group into subgroups, when trait is normally distributed.

6.3.5.1 Cases within given score limits:

Generally three types of cases are encountered under this type—percentage of cases below a score point; percentage of cases above a score point; and percentage of cases within two score limits.

Example

Given a normal distribution of 500 scores with $M=40$, and $\sigma=8$, what percentage of cases lie (a) between scores of 36 and 42; (b) above a score of 48; and (c) below a score of 52.

Solution

(a)

<i>Score limits</i>	<i>Equivalent z</i>	<i>Normal curve table reading</i>	<i>% Area or % cases</i>
36	$(36-40)/8 = -.5$.1915	19.15
42	$(42-40)/8 = +.25$.0987	9.87
			29.02

$$\text{Cases out of } N = \% \text{ area} \times N/100 \quad (6.14)$$

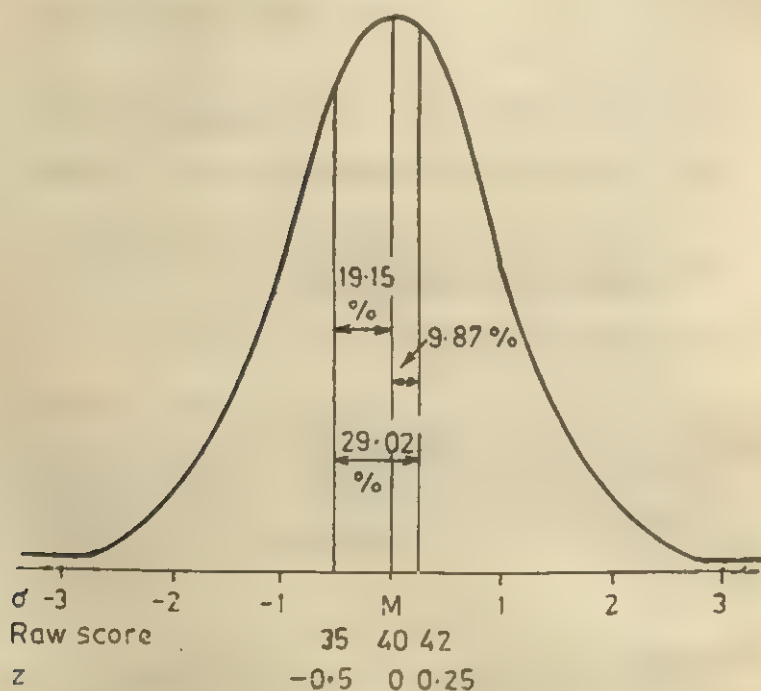


Fig. 6.2. Calculation of Area when z Limits fall on both sides of the Mean.

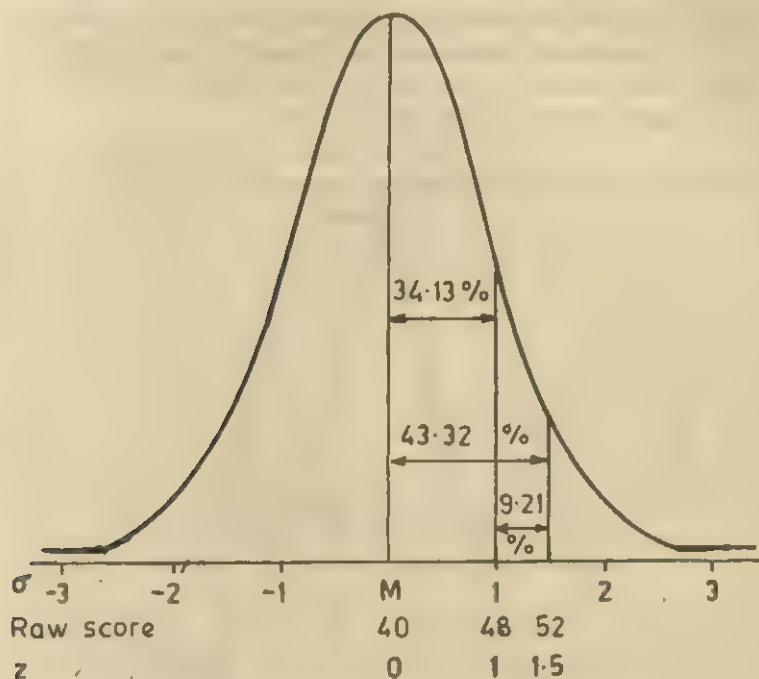


Fig. 6.3. Calculation of Area when z Limits fall on one side of the Mean.

Hence cases out of $500 = 29.02 \times \frac{500}{100} = 145.10$ or 145 cases.

- (i) The two score limits have been converted into equivalent z's by using the formula $(X-M)/\sigma$ and algebraic signs also noted (see Col. 2).
- (ii) Normal curve area tables consulted to find out readings equivalent to z's (see Col. 3).
- (iii) Since the two z values have different signs, the one with a negative sign falls .5 σ units below the mean and the other falls .25 σ units above the mean. Hence, the two percentages of the area are to be added.
- (iv) The percentage of area can be converted into number of cases by using formula (6.14).
- (v) In cases where the two values of z have the same sign or fall in the same half (i.e. only in the lower half, or

only in the upper half), the smaller area is to be subtracted from the larger one. See Figure 6.3 which has been constructed on the basis of score limits of 52 and 48, both of which fall in the upper half.

(b) Percentage of area above a score of 48

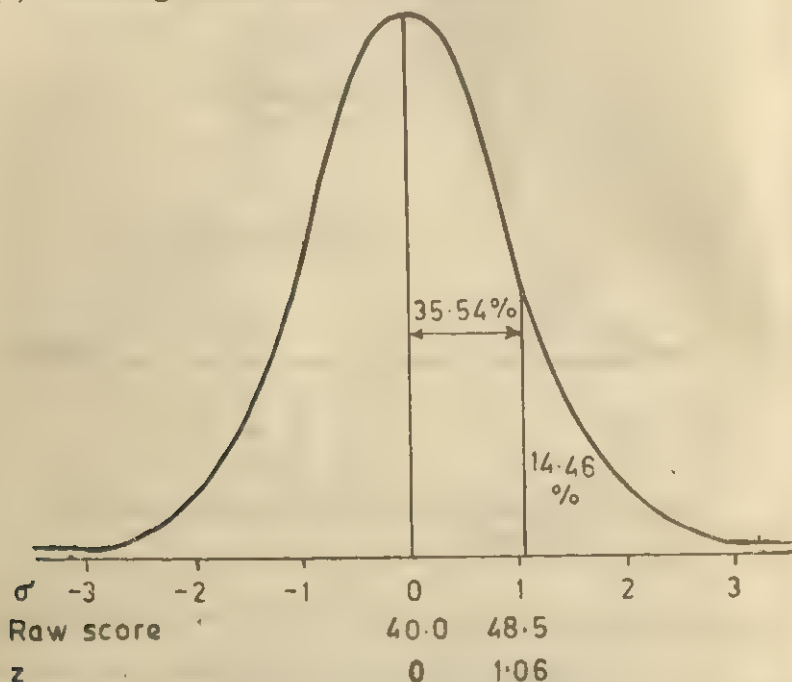


Fig. 6.4. Calculation of Area above a Given Score.

Raw score	Exact upper limit of score	z Score	Table reading	%Area
48	48.5	$(48.5 - 40) / 8 = 1.06$ (Rounded)	.3554	35.54

Since the total area in the upper half is 50 per cent, and the area between Mean and 1.06 z is 35.54 per cent, the area above a z of 1.06 is obtained by subtracting 35.54 per cent from 50.00 per cent. Hence the area above a score of 48 is equal

to 50.00 per cent - 35.54 per cent = 14.46 per cent and the corresponding number of cases out of 500 are $= 14.46 \times 500, 100 = 72.30$ or 72 cases.

(c) Percentage of area below a score of 52

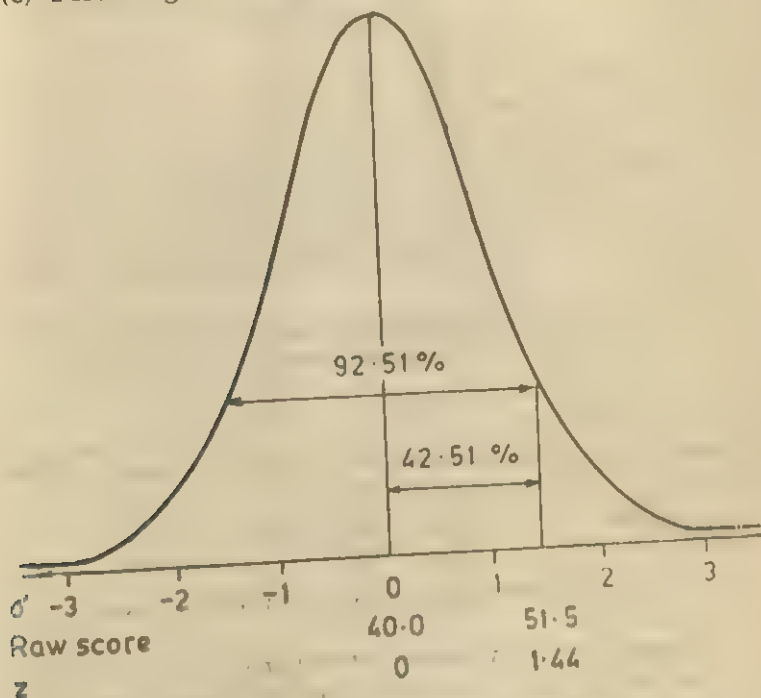


Fig. 6.5. Calculation of Area below a Given Score.

Raw score	Exact lower limit	z score	Table reading	% Area
52	51.5	$(51.5 - 40.00) / 8 = +1.44$ (Rounded)	.4251	42.51

The area below a z of 1.44 would include the whole of the lower half i.e., 50 per cent, and 42.51 per cent i.e., from mean to $z=1.44$. Hence, by adding, we obtain $50.00 + 42.51 = 92.51$ per cent or $92.51 \times \frac{500}{100} = 462.55$ or 463 cases out of 500.

6.3.5.2 Limits of Scores which include a given percentage

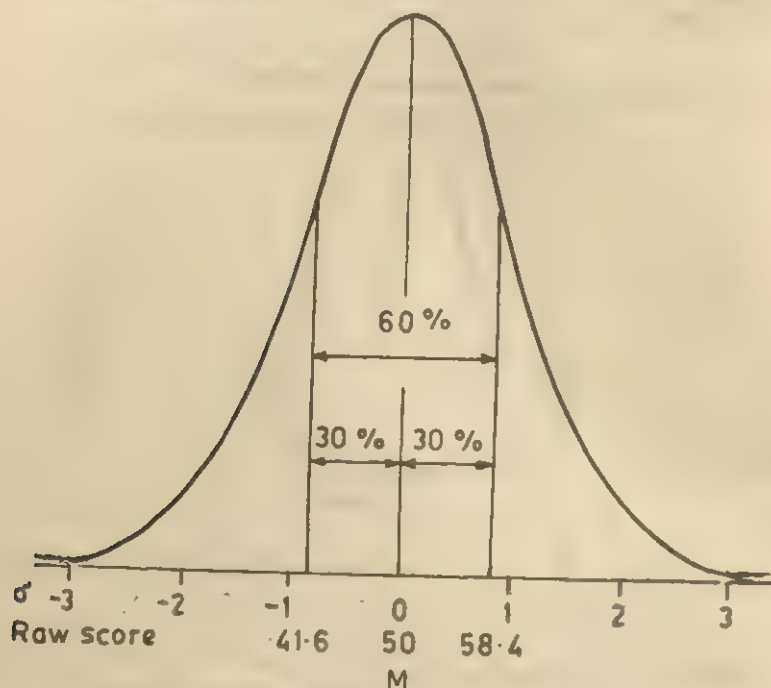


Fig. 6.6. Score Limits Equivalent to Middle 60% Cases.

Example: Given a normal distribution of 500 scores with a mean of 50 and σ of 10. What score limits would include the middle 60 per cent cases?

Solution

To obtain the middle 60 per cent cases, one has to find out the score limits which combine 30 per cent cases below the mean and 30 per cent cases above the mean. Hence, the normal curve area tables are to be consulted in a reverse manner. Locate the 30 per cent area in Table A and read z equivalent to it. Then use the following formula to obtain the score limits:

$$\text{Raw score} = M + z \times \sigma \quad (6.15)$$

The procedure is illustrated below :

<i>Area percentage</i>	<i>Equivalent z with sign</i>	$z \times \sigma$	<i>Raw score $M + z\sigma$</i>
30 % below M	-.84	$.84 \times 10 =$ -8.4	$50 + (-8.4)$ =41.6
30 % above M	+.84	$.84 \times 10 =$ 8.4	$50 + 8.4$ =58.4

The z values below the mean will carry minus signs and those above the mean, plus signs. Thus, the score limits of 41.6 - 58.4 include the middle 60 per cent cases. In a similar manner other score limits can be calculated.

6.3.5.3 Comparison of two distributions in terms of 'Overlapping'

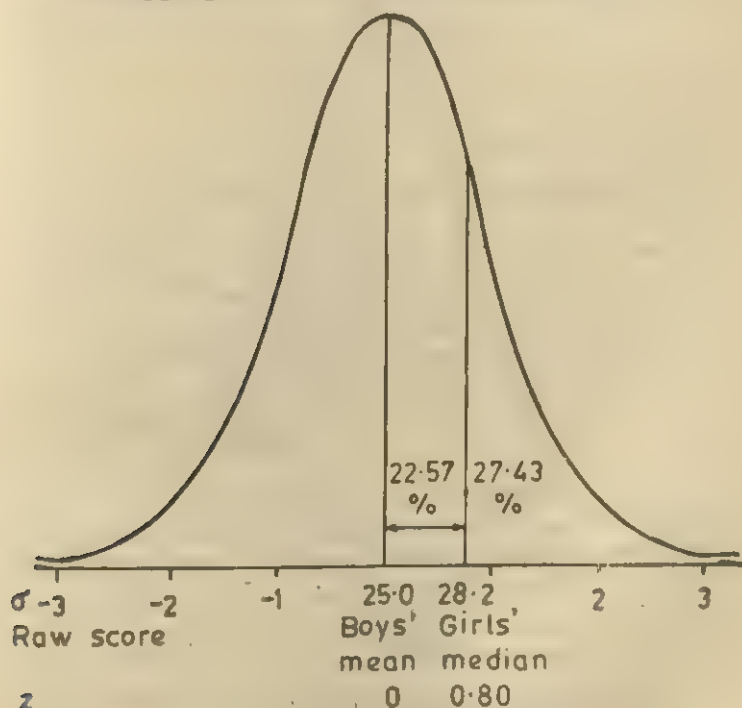


Fig. 6.7. Percentage of Cases Exceeding a Particular Score.

Example: Given two normal distributions of scores made on a numerical ability test by 200 boys and 300 girls. The boys' mean score is 25 with a σ of 4. The girls' mean score is 28 with a σ of 6. The medians are: boys, 25.2 and girls, 28.2. What percentage of boys exceeds the median of the girls' distribution?

Solution

The girls' median is $28.2 - 25.0$ or 3.2 score units above the boys' mean. Dividing 3.2 by 4 (the σ of boys' distribution), we find that girls median is .80 σ above the mean of the boys' distribution. From the normal curve area table, we find that 22.57 per cent of the normal distribution lies between the mean and .60 σ . Hence, 27.43 per cent of the boys (50 per cent 22.57 per cent) exceed the girls' median.

6.3.5.4 Determination of relative difficulty of test items

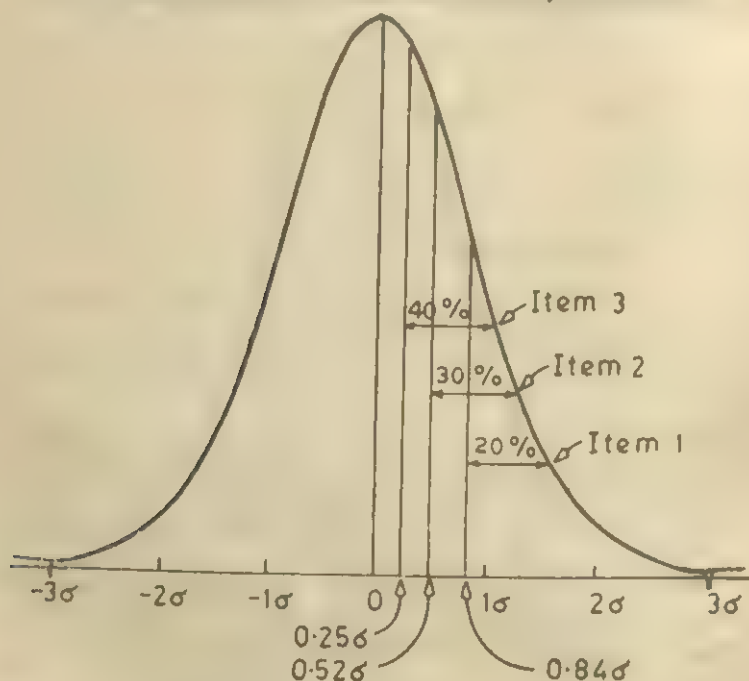


Fig. 6.8. Comparison of Relative Difficulty Value of test Items Based on Sigma Differences.

Example: A test item is answered correctly by 20 per cent of a large unselected group; a second item, solved correctly by 30 per cent of the same group; and a third problem solved by 40 per cent. Assuming normal distribution of capacity to solve test problems, what is the relative difficulty of items 1, 2 and 3?

Solution

First of all, we shall find out the cut off points of the baseline of the curve, which shows top 20 per cent, top 30 per cent, and top 40 per cent of the cases. Since all the distances are to be taken from the mean, the areas to be looked up in the Normal curve table are $(50-20)=30$ per cent; $(50-30)=20$ per cent; and $(50-40)=10$ per cent.

Now find out the z 's equivalent of these three percentages:

Item	Passed by	z	z difference
1	20%	.84 σ	—
2	30%	.52 σ	.32 σ
3	40%	.25 σ	.27 σ

We may now compare the three items on the difficulty level based on z 's. Item 1 has a difficulty value of .32 σ higher than item 2 and Item 2 has a difficulty value of .27 σ higher than item 3. Thus, the difference in difficulty value of item 2 and 3 is about $3/4$ of the difference in difficulty value of item 1 and 2. Thus the difference in percentages (which is equal to 10 per cent in each case here) is not as good an index of differences in difficulty as the σ difference is.

6.3.5.5 Division of a group into sub-groups

Example: Given a group of 500 College students who have been administered a general mental ability test. We wish to classify our group into five sub-groups A, B, C, D and E according to ability, the range of ability being equal in each sub-group. It is assumed that the trait is normally distributed in the population. Calculate the number of students that can be placed in groups A, B, C, D and E.

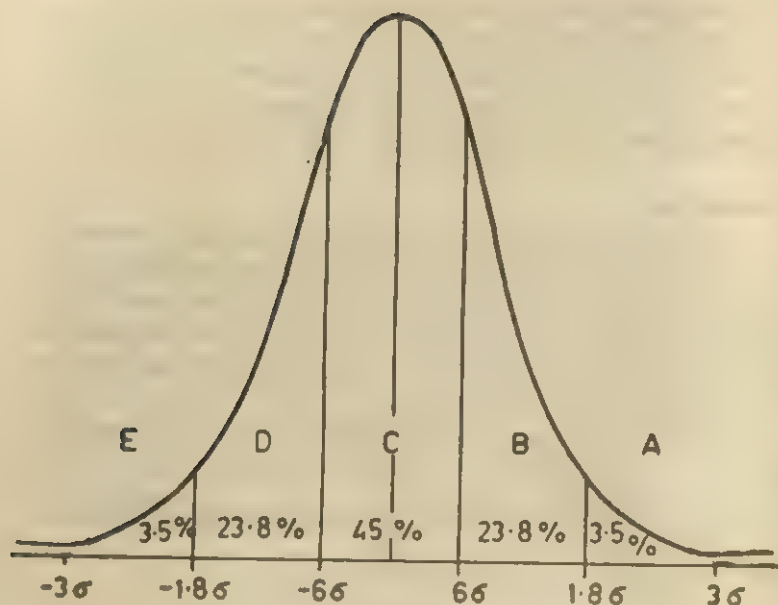


Fig. 6.9. Classification of a Group into Sub-Groups.

Solution

The baseline of a normal curve is considered to extend from -3σ to 3σ that is over a range of 6σ . Dividing this range by 5 (the number of sub-groups) we obtain 1.2σ . Each group is to be allotted an equal extent of 1.2σ on the baseline. The five intervals, thus formed, have been shown in Figure 6.9.

Group A covers the upper 1.2σ ; group B the next 1.2σ ; group C has $.6\sigma$ to the right and $.6\sigma$ to the left of the mean; Groups D and E occupy the same relative positions in the lower half of the curve that B and A occupy in the upper half. The next step is to find the percentage of the area that falls in the upper and lower limits of each sub-group. Calculation of these percentages is shown below:

The number of cases in each sub-group as shown in the last column have been obtained by multiplying the percentage of area within a sub-group by $500/100$ or 5.

In a similar manner, the group can be classified into 6 sub-groups by dividing the total range of 6σ by 6 and partitioning

TABLE 6.8
Calculation of Percent Area and Number of Cases in each sub-group

Sub-group	Upper limit		Lower limit		n% in sub-group	Cases out of 500
	σ	%	σ	%		
A	+3.0	49.86	+1.8	46.41	49.86-46.41= 3.45	18
B	+1.8	46.41	+ .6	22.57	46.41-22.57=23.84	119
C	+ .6	22.57	- .6	22.57	22.57+22.57=45.14	226
D	- .6	22.57	-1.8	46.41	46.41-22.57=23.84	119
E	-1.8	46.41	-3.0	49.86	49.86-46.41= 3.45	18

the baseline into 6 intervals of one σ each. For classification into four sub-groups, the baseline is to be partitioned into 4 equal intervals of 1.5σ each ($6 \sigma/4 = 1.5 \sigma$).

6.3.6. Importance of the Normal Distribution

The normal distribution is by far the most used distribution in inferential statistics. There are some very important reasons for the same.

1. A good fit to many phenomena

Sufficient evidence has accumulated to show that the normal distribution provides a good fit to describe the frequency of occurrence of many variable facts in biological statistics (sex ratio in births in a country over a number of years); in anthropometrical data (height, weight, etc.); in social and economic data (rates of births, marriages or deaths); in psychological measurements (intelligence, perception span reaction time, educational test scores); and in errors of observation (chance errors which cause observations of heights and weights to deviate above or below their true value).

2. It may be convenient, on mathematical grounds alone, to assume a normally distributed population.

The normal function has important mathematical properties shared by no other theoretical distribution. Assuming a normal distribution gives the statistician an extremely rich set of mathematical consequences that he can use in developing methods.

3. The normal distribution can be used as a good approximation to a number of other theoretical distributions like the binomial and can allow the calculation of probabilities that are either laborious or impossible to work out exactly.

4. There is a very intimate connection between the size of sample and the extent to which a sampling distribution approaches the normal form.

The central limit theorem states that many sampling distributions based on large N follow the normal curve even though the population distribution is definitely non-normal. This is one of the most remarkable and useful principles used in inferential statistics.

6.4. Divergence From Normality

6.4.1. Skewness

When a distribution of scores is not symmetrical, it is said to be asymmetrical or skewed. By skewness, then, we mean the degree of its departure from symmetry. The frequency distribution of a set of scores is called symmetrical about the mean if the number of frequencies at any point on the upper side of the mean is exactly the same at a point equidistant from the mean on the lower side. Measures of skewness indicate two things—the magnitude of skewness and the direction of skewness. The symmetry of a curve is disturbed by the bunching of scores on one side of the central tendency or to the trailing out of scores in one direction from the central tendency.

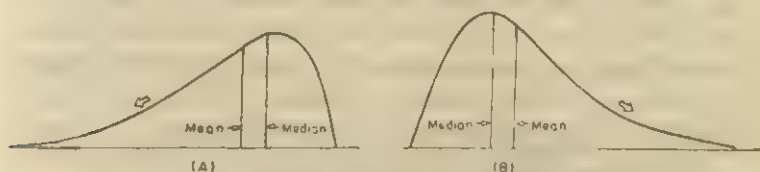


Fig 6.10. (A) Negative Skewness : to the Left.

(B) Positive Skewness : to the Right.

Fig. 6.10(A) is *negatively skewed* or *skewed to the left* because the scores tend to trail off to the left or the negative end of the curve. Fig. 6.10(B) is *positively skewed* or *skewed to the right* because the scores tend to trail off to the right or the positive end of the curve.

A simple method of detecting the direction of skewness by inspection is by looking at the tails of the distribution. The simple rule is: if the longer tail of the distribution is towards the higher values or upper side, the skewness is positive; if the longer tail is towards the lower values or lower side, the skewness is negative. A measure of skewness based on mean and median is given by the formula:

$$SK = \frac{3(\text{mean} - \text{median})}{\sigma} \quad (6.16)$$

Another measure of skewness based on percentiles is given by the formula:

$$SK = \frac{(P_{90} + P_{10})}{2} - P_{50} \quad (6.17)$$

The two measures are not mathematically equivalent. A normal curve has $SK = 0$. Deviations from normality can be in negative and positive directions leading to negatively skewed and positively skewed distributions respectively.

6.4.2 Kurtosis

The Kurtosis of a distribution refers to its 'curvedness' or 'peakedness'. Two distributions may have the same mean and the same variance and may be equally skewed, but one of them may be more peaked than the other. The peakedness is based on the degree of concentration of the scores near the central tendency. A normal curve is *mesokurtic* or having medium kurtosis. A distribution having a high concentration of scores near the central tendency and high tails as compared to a normal distribution with the same standard deviation is called *leptokurtic* (lepto means slender or narrow). A distribution having low concentration of scores in the neighbourhood of the central tendency and low tails as compared to a normal distribution with the same standard deviation is called *platykurtic* (platy means flat, broad or wide).

Fig. 6.11 shows three curves depicting mesokurtosis, leptokurtosis and platykurtosis.

Several rough methods can be used for judging whether a distribution lacks normal symmetry or peakedness. It may ordinarily be detected by inspection of the frequency polygon. However, apparent peakedness may result from choice of dimensions for the polygon. Another way of detecting kurtosis is by seeing whether the percentages of cases in various quartile or standard deviation intervals included normal percentages of cases. If these percentages are not normal, the distribution is either skewed or abnormally peaked.

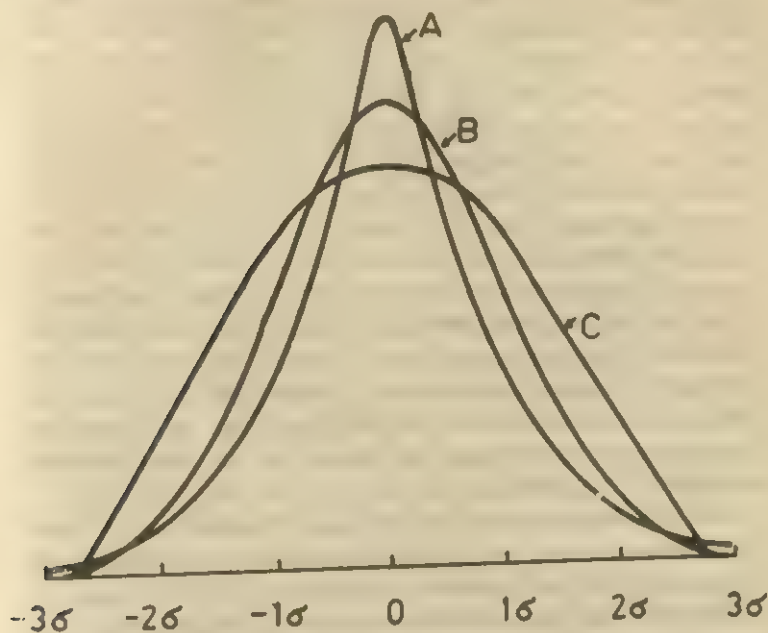


Fig. 6.11. (A) Leptokurtic; (B) Normal or Mesokurtic;
(C) Platykurtic Curves.

Kurtosis can be measured by the following formula based on percentiles:

$$KU = \frac{Q}{(P_{90} - P_{10})} \quad (6.18)$$

where, Q = quartile deviation,

P_{90} and P_{10} = 90th and 10th percentiles respectively.

A normal distribution has $KU = .263$. If KU is less than .263, the distribution is leptokurtic; and if KU is greater than .263, the distribution is platykurtic.

6.5. Measures of Skewness and Kurtosis based on Moments Methods

In mechanics, the term "moment" is used to denote a measure of the tendency of a force to cause rotation of an

object about a point. The strength of the tendency depends upon two things—the amount of the force and the distance from the point at which the force acts. Hence, a moment is the product of force times distance ($M = F \times D$). There may be moments applying the force in one direction and others applying it in the other direction. When the sum of the moments tending to cause rotation in one direction is equal to the sum of the moments tending to cause rotation in the opposite direction, the object is in balance. In a statistical series, an item may be considered as a unit force acting at a distance x from the arithmetic mean i.e., as a moment of force. Since the sum of negative deviations from the mean is equal to the sum of positive deviations, the mean is analogous to a point of balance. In statistics, the algebraic sum of the distance or deviations from the mean divided by N is called the first moment of the series.

The second moment of the series can be obtained if deviations from the mean are squared, summed and divided by N . The third and fourth moments are based respectively upon the third and fourth powers of the deviations. Thus, the first four moments about the mean can be defined as follows:

$$\mu_1 = \frac{\sum x}{N} = 0 \quad (6.19)$$

$$\mu_2 = \frac{\sum x^2}{N} = \sigma^2 \quad (6.20)$$

$$\mu_3 = \frac{\sum x^3}{N} \quad (6.21)$$

$$\mu_4 = \frac{\sum x^4}{N} \quad (6.22)$$

where, $\mu_1, \mu_2, \mu_3, \mu_4$ are first, second, third and fourth moments respectively, and $x = X - M$ and the deviations are taken from the actual mean. When data are grouped in terms of class intervals of the constant size i and frequencies, the second, third and fourth moments about the mean can be calculated by using the following formulas:

$$\mu_2 = i^2 \left[\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N} \right)^2 \right] = \sigma^2 \quad (6.23)$$

$$\mu_3 = i^3 \left[\frac{\sum fd^3}{N} - 3 \left(\frac{\sum fd^2}{N} \right) \left(\frac{\sum fd}{N} \right) + 2 \left(\frac{\sum fd}{N} \right)^3 \right] \quad (6.24)$$

$$\mu_4 = i^4 \left[\frac{\sum fd^4}{N} - 4 \left(\frac{\sum fd^3}{N} \right) \left(\frac{\sum fd}{N} \right) + 6 \left(\frac{\sum fd^2}{N} \right) \left(\frac{\sum fd}{N} \right)^2 - 3 \left(\frac{\sum fd}{N} \right)^4 \right] \quad (6.25)$$

It may be noted that the calculation of μ_2 is exactly alike that of calculation of variance by the coded method. The initial steps in the calculation of other moments also follow the same method with the difference that these go up to $\sum fd^3$ in case of μ_3 and upto $\sum fd^4$ in case of μ_4 . The value of $\sum fd^3$ can be obtained by multiplying each fd^2 by corresponding d and summing over all such values thus obtained. The value of $\sum fd^4$ is obtained by multiplying the entries in fd^3 further by d and obtaining the column sum. These values may then be divided by N separately and substituted in the formulas given above. There is hardly any need to substitute the value of i (class interval) and express the moments in the original units of measurements. When the moments are substituted in formulas 6.26 and 6.27, the i 's cancel out. Necessary attention regarding the correct substitution and multiplication signs should be given to avoid errors in computing moments.

After the moments have been obtained, skewness and kurtosis can be computed by using the formulas given below:

$$a_3 = \frac{\mu_3}{\mu_2 \sqrt{\mu_2}} \quad (6.26)$$

$$a_4 = \frac{\mu_4}{\mu_2^2} \quad (6.27)$$

in which, a_3 and a_4 = skewness and kurtosis, respectively,
 μ_2 , μ_3 and μ_4 = second, third and fourth moments, respectively.

Interpretation: For the purpose of interpretation of a_3 and a_4 , the following quick guide may be used.

value		Interpretation
$a_3 = 0$		normal
$a_3 < 0$	(negative value)	negative skewness
$a_3 > 0$	(positive value)	positive skewness
$a_4 = 3$		normal
$a_4 < 3$		platykurtosis
$a_4 > 3$		leptokurtosis

In case of a_3 , the greater the departure from 0, the greater the negative or positive skewness. In case of a_4 , the greater the departure from 3, the greater the platykurtosis or leptokurtosis.

6.5.1 Significance of the Measures of Skewness and Kurtosis

The values of a_3 and a_4 obtained for any distribution are compared with the values of a_3 and a_4 found in normal distributions and significance of the difference determined. The questions previously are: does the obtained value of a_3 differ significantly from zero? If yes, at what level of confidence? Does the obtained value of a_4 differ significantly from 3? If yes, at what level of confidence?

Byron S. Pearson has devised table of 0.10 and 0.02 limits of a_3 when based on samples drawn from a normal population. Another table showing the upper and lower 0.1 and 0.02 limits of a_4 when based on random samples from a normal distribution has also been prepared. These tables are given in Appendix II and I.

6.6 Importance of Measures of Skewness and Kurtosis

The measures of skewness and kurtosis are of interest to the researchers and statisticians for a number of reasons and uses which these can be put to. Some more important of them are discussed below in brief.

(1) Measures of skewness and kurtosis together with measures of central tendency and variability, and the number of cases, N , convey all the information ordinarily needed to understand and interpret any distribution.

(2) Measure of skewness and kurtosis indicate the extent of 'non-normal' variation in a set of data. Since a good number of parametric tests, parametric correlation, at some point of their application require symmetry in departure from normality, the measures of skewness and kurtosis can be helpful in identifying whether the assumption of symmetry is fulfilled by the experimental data. In general, the use of a post-hoc statistical test (the t -test) of the difference in means of that control and test statistic may not be sufficient to determine, especially, existence of outliers, trends, and variability.

It can further be argued that these threshold quantities indicate departures from normality and hence provide a means for case and control selection. In other words, exposure to an agent is not normal, and hence "abnormal" exposures are found, and a search is warranted. At the same time, the normality of the way the model is fitted to the data is assumed, and the departure from the normality of the data is the pointer to the departure from normal peakedness.

(d) Another important method with the same goal of newness and variety is *random search* (see Figure 1 for a sketch of the shape of the search space). Here, the function is additionally formulated to support not only *best-of-known* measures of center, but also *best-of-random*. This is due to the shape of the search space, which is *not* the average value.

(4) Education and personnel management are the two most important areas for the future. The future of the country depends on the quality of the human resources. The government should invest more in education and training, and improve the quality of the workforce. The government should also improve the management of the public sector, and improve the quality of the public services.

Since the main purpose of secondary measures of education and health is to aid in the diagnosis of developmental distribution and to facilitate comparisons between two or more distributions, it would be desirable to have standard tables certain points on the tail of the curve. The frequent occurrence of some critical points in the frequency of educational variables may result from a continuous distribution of a skewed distribution and may be utilized. Most of the instruments in possession of the Bureau of

measurements involve unequal units decided upon arbitrarily and sometimes accidentally. Generally, the score a person gets on a test depends upon the number of items checked or done correctly. Since the number of items correct is an artifact of the difficulty level of the items the shape of the distribution is largely determined by the latter.

With items of medium difficulty, the scale is likely to yield a symmetrical distribution when applied to a group, if the items are very easy, the scores will pile up toward the top thus leading to negative skewness; if the items are very difficult, the concentration of scores toward the bottom will occur, thus generating a positive skewness.

Ordinarily, it is not necessary to compute measures of skewness and kurtosis unless there are indications to the same. However, the nature of the investigation, the size of the sample and the nature of the variable under study are important factors to be considered in determining whether these measures should be computed or not. If the size of the sample is less than 100, these measures may not give very reliable results.

Exercises for Practice

- 6.1 In two throws of a coin what is the probability of throwing :
- (a) both heads
 - (b) both tails
 - (c) at least one head
 - (d) at least one tail
 - (e) one head or one tail.
- 6.2 If three coins are tossed, what is the possibility of obtaining :
- (a) At least two heads
 - (b) at least one head
 - (c) all three heads
 - (d) all three tails
 - (e) Exactly one head.

- 6.3 If the probability of answering a question correctly is 5 times the probability of answering it incorrectly, what is the probability of answering it correctly ?
- 6.4 In a six-item test each item is scored right or wrong. If a student answers each item by guessing alone, (a) How many different outcomes are possible ? (b) State the binomial. (c) Describe the different types of results. (d) What is the probability of getting (i) 4 or more items correct ? (ii) All the 6 items correct. (iii) Exactly 3 items correct. (e) If 64 students attempted the test by guessing, how many students would obtain the situation d(i); d(ii); and d(iii) above.
- 6.5 If the probability of having a boy or a girl is equal, out of 64 families with four children each, how many would you expect to have (a) exactly one boy, (b) exactly three boys, (c) at least one boy, and (d) all girls.
- 6.6 If the probability of selecting an extrovert is .40. Out of a sample of 5 individuals selected independently, what is the exact probability that three of the individuals selected are extroverts and the remaining two are not ?
- 6.7 (a) Define a Normal Distribution ? What are its principal properties ?
(b) What is a unit normal curve ?
- 6.8 In a group of 500 students based on normal distribution, $M=40$; $SD=8$, find out:
(a) Number of students between scores of 25 and 35.
(b) Values of P_{75} and P_{50} .
(c) If the group is divided into six subgroups on the basis of equal spread of ability, what will be the number in each subgroup ?
- 6.9 Define the following :
(a) Skewness,
(b) Kurtosis,



- (c) Platykurtosis,
 - (d) Leptokurtosis,
 - (e) Negative and positive skewness.
- 6.10 Calculate, by moments method, the value of S_k and K_u from the following set of scores:
- Scores:* 5, 4, 8, 8, 5.

CHAPTER 7

CORRELATIONAL TECHNIQUES

7.1 The Concept

So far, we have looked into the distribution of a single variable at a time. Very often, however, there is interest in examining as to how variations in one variable are associated with or related with the variations in another variable. This is called a bivariate situation. For this purpose, we need an index of association or relationship between the two variables. This is known as *Coefficient of Correlation*. *A coefficient of correlation is a single number that tells us to what extent two variables or things are related and to what extent variations in one variable go with variations with the other.* Whenever two measurements for the same individual can be paired for all the individuals in a group, the degree of relationship between the paired scores is called "correlation".

For example, a teacher finds that the students having high I.Q.'s have secured high marks and students having low I.Q.'s have secured low marks on a test of academic achievement. It shows a definite trend of relationship in the variations of I.Q.'s and marks. Let us examine the three sets of X and Y scores given in Table 7.1 for 10 children. Scores for the X variable have been ordered from the highest to the lowest and remain uniformly so in all the three A, B and C situations. Data for the Y variable are arranged in three different ways corresponding to the column headings, A, B and C.

The data in column A show that each person attains the same score on both variables. The child with the highest X score has also the highest Y score; conversely the child with the

TABLE 7.1

Paired Scores for Three Levels of Correlation

Persons	A		B		C	
	X	Y	X	Y	X	Y
1	20	20	20	11	20	19
2	19	19	19	12	19	14
3	18	18	18	13	18	13
4	17	17	17	14	17	11
5	16	16	16	15	16	20
6	15	15	15	16	15	12
7	14	14	14	17	14	17
8	13	13	13	18	13	18
9	12	12	12	19	12	16
10	11	11	11	20	11	15
Interpretation	Perfect Positive $r = +1.00$		Perfect negative $r = -1.00$		Zero correlation $r = 0$	

lowest score on X has the lowest score on Y. Each X score corresponds exactly with each Y scores. This leads to a perfect relationship. Since an increase in the value of X scores results in a corresponding increase in the Y scores, the direction of relationship is positive. The scatter plot for set A shown in Figure 7.1 represents the placement of scores along a straight line which runs from the lower left hand corner to the upper right hand corner. The correlation coefficient, r , in this case is 1.00.

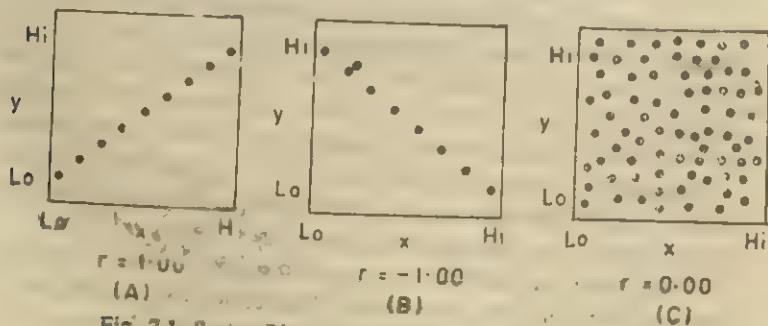


Fig. 7.1. Scatter Plots for Three Levels of Correlation.

The scores in Column B of Table 7.1 show a trend which is exactly opposite to the trend of set A. The person obtaining the highest score on the X variable has the lowest score on the Y variable. Conversely, the lowest score on the X variable corresponds to the highest score on the Y variable. The positions of other scores have also been reversed in a perfectly systematic way. This is a case of a perfect or maximal correlation but with a negative direction. The scatter plot for set B in Figure 7.1 represents the placement of scores along a straight line which runs from the lower right hand corner to the upper left hand corner. The correlation coefficient in this case -1.00 . Compare the direction of the straight lines of scatter plots A and B. What is the difference?

The scores in column C of Table 7.1 show that it was difficult to find any particular trend of association between the variations in the scores of X and Y. The correlation is essentially non-existent. The scatter plot for set C in Figure 7.1 shows that the scores fall all over the surface of the graph in such a way that change or variation in one variable is unrelated with the other variable. Hence value of $r=0.00$.

7.2 The Product Moment Correlation. r

Karl Pearson's Product Moment Coefficient of Correlation can be computed by using the definitional formula:

$$r = \frac{\sum z_x z_y}{N} \quad (7.1)$$

The student will recall that z or standard scores can be obtained by dividing the deviation scores by standard deviation. Hence, using the paired z scores, the product moment correlation can be defined as the average product of the paired z scores. It can be shown algebraically that the maximum value of the term $\sum z_x z_y$ is attained when, for each pair of z scores, $z_x = z_y$. It can also be shown that when the z scores for each pair are identical, the sum of their products equals N . Using standard scores for the calculation of correlation coefficient involves a tedious process. However, several formulas exist which have been derived from the basic definitional formula and involve less labour in the computation of r . The process of calculation of r is explained below:

TABLE 7.2

Calculation of Product Moment r by two different Formulas

Raw Score Method				
X	Y	X^2	Y^2	XY
10	11	100	121	110
8	7	64	49	56
6	2	36	4	12
4	6	16	36	24
2	4	4	16	8
30	30	220	226	210
ΣX	ΣY	ΣX^2	ΣY^2	ΣXY

Formula:

$$\begin{aligned}
 r &= \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N \Sigma X^2 - (\Sigma X)^2][N \Sigma Y^2 - (\Sigma Y)^2]}} \quad (7.2) \\
 &= \frac{5(210) - (30)(30)}{\sqrt{[5(220) - (30)^2][5(226) - (30)^2]}} \\
 &= \frac{1050 - 900}{\sqrt{(1100 - 900)(1130 - 900)}}
 \end{aligned}$$

Deviation Score Method				
x	y	x^2	y^2	xy
4	5	16	25	20
2	1	4	1	2
0	-4	0	16	0
-2	0	4	0	0
-4	-2	16	4	8
$M_x - 6$	$M_y - 6$	Σx^2	Σy^2	Σxy

Formula:

$$\begin{aligned}
 r &= \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \quad (7.3) \\
 &= \frac{30}{\sqrt{(40)(46)}} \\
 &= \frac{30}{\sqrt{429}}
 \end{aligned}$$

$$= \frac{150}{\sqrt{200 \times 230}}$$

$$r = .7$$

$$r = .7$$

Steps

1. Calculate squares of each X and Y score as shown in Cols. (3) and (4)
2. Calculate products of X and Y by multiplying the value of Cols. (1) and (2)
3. Sum up all columns.
4. Insert the values in the formula.

Steps

1. Calculate Means of X and Y scores
2. Obtain deviation scores x and y as follows:
 $x = X - M_x; y = Y - M_y$
3. Obtain squares of deviation scores.
4. Multiply Cols. (6) and (7) to obtain xy.
5. Sum up all columns
6. Insert the values in the formula.

7.3 Some Other Formulas

In Table 7.2, two different formulas have been used for the calculation of r . Both the formulas are mathematically equivalent and give the same results. The formula based on the raw scores is more convenient when a calculating machine is available or the analysis is to be done on a computer. The formula using deviation scores is more useful when N is small and means of X and Y scores are whole numbers. Another formula based on raw scores and mathematically equivalent to the two formulas mentioned in Table 7.2 is

$$r = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)}{N} \div \sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N} \right] \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N} \right]} \quad (7.4)$$

(Coefficient of correlation from raw scores)

Another formula using raw scores and mean is as follows:

$$r = \frac{\Sigma XY - N M_x M_y}{\sqrt{(\Sigma X^2 - N M_x^2)(\Sigma Y^2 - N M_y^2)}} \quad (7.5)$$

(Coefficient of correlation using raw scores and means)

where M_x and M_y are means of X and Y scores, respectively. Difference formula may also be used for the calculation of r . The same is given below:

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma d^2}{2\sqrt{\Sigma x^2 \times \Sigma y^2}} \quad (7.6)$$

(Coefficient of correlation using difference formula, deviations from the means of the distributions)

where $\Sigma d^2 = \Sigma (x-y)^2$; other symbols are the same as used above. The difference formula does not require the calculation of cross products (xy 's). The student may try the difference formula on scores given in Table 7.2. All values, except Σd^2 , have already been presented there. Another variation of the difference formula which is often useful in machine calculation is also given here. This version uses raw or obtained scores instead of deviation scores.

$$r = \frac{N[\Sigma X' - \Sigma Y^2 - \Sigma(X-Y)^2 - (\Sigma X)(\Sigma Y)]}{2\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \quad (7.7)$$

(Coefficient of correlation by difference formula based on raw or obtained scores)

in which $\Sigma(X-Y)^2$ is the sum of the squared difference between X and Y sets of scores.

The student may use this formula on data given in Tab¹ 7.2. All calculations, except $\Sigma(X-Y)^2$, are already available therein.

7.4 Spearman's Rank-Difference Correlation Coefficient (rho)

In case when distributions of scores are markedly skewed, measurements made with an interval or ratio scale can be transformed to ranks before the correlation is computed. Sometimes only ordinal scale data in the form of ranks are available and the calculation of product moment r is not possible. In such situations, Pearson's rank-order correlation coefficient, rho, can be calculated. It is also useful when N is very small. The assumption of normal distribution of the characteristic in the population is not required. The procedure of calculation of rho is shown in Table 7.3.

7.4.1 Calculation of rho when no ties exist

TABLE 7.3

Rank-Difference Coefficient of Correlation (Case of no ties)

Student	Score on Test I X	Score on Test II Y	Rank on Test I R ₁	Rank on Test II R ₂	Difference between ranks D	Difference squared D ²
A	8	4	2	5	-3	9
B	7	7	3	3	0	0
C	9	6	1	4	-3	9
D	5	8	4	2	2	4
E	1	10	5	1	4	16
N=5						$\Sigma D^2=38$

Formula:

$$\begin{aligned}
 \rho &= 1 - \frac{6\sum D^2}{N(N^2-1)} & (7.8) \\
 &= 1 - \frac{6 \times 38}{5(5^2-1)} \\
 &= 1 - \frac{228}{120} \\
 &= 1 - 1.9 \\
 &= -0.9
 \end{aligned}$$

Interpretation: Relationship between X and Y is very high and inverse.

In Table 7.3, students have been listed in Col. (1); scores on Test I and Test II, in Cols. (2) and (3). There are no ties in the scores, and ranking is simple. It can be accomplished by assigning a rank of 1 to the highest score, a rank of 2 to the next highest, and so on, till the lowest score gets a rank equal to N. Take one set of scores at a time and finish the ranking. Then take up the second set. In Table 7.3, student C gets the highest X score of 9 and hence obtains a rank of 1 (see Col. 4); Student A with the next highest X score of 8 gets the next rank of 2; student B, a rank of 3; student D, a rank of 4; and student E, with the lowest score, gets the last rank of 5 which is equal to N. The same procedure is repeated for ranking the students on Test II and the ranks written in Col. (5). In column (6), differences of ranks for each student are mentioned. For example, for student A; $D = R_1 - R_2$ or $2 - 5 = -3$. Col. (7) shows each rank difference squared. Col. (7) is summed up to obtain $\sum D^2$ which is 38 in our example. The number of students in this example is five, hence $N=5$. The numerical values 1 and 6 in the formula are constants and remain the same in all cases. The values of $\sum D^2$ and N are inserted in the formula and the value of rho obtained. Necessary precaution should be taken regarding the sign of the value of rho when it is negative lest it is omitted inadvertently. In our example, $\rho = -0.9$ a value which is quite high and shows an inverse relationship between X and Y.

7.4.2 Calculation of rho when tied ranks exist

Sometimes, two or more persons obtain the same score or have the same age, years of service or some other numerical value. This introduces ties in the scores which are then reflected in their ranks also. The procedure of assigning ranks in such situations will be described with reference to the example solved in Table 7.4. Column (1) shows student serial number in terms of letters. In Cols. (2) and (3) each students' scores on Test I and Test II have been given. For ranking, one set of scores, say X, is taken at a time. Ranks assigned to scores on Test I have been shown in Col. (4). Student F with the highest score of 20 has been given a rank of 1; and student E whose score is the next highest, a rank of 2. In this way, the ranking has been completed till student I got a rank of 10. Students A and G have similar scores of 10 each and they possess 6th and 7th positions. Instead of assigning either 6 or 7 to both of them, the average of the two positions, i.e. $(6+7)/2=6.5$ has been assigned to each of them.

The same procedure has been followed in respect of scores on Test II. In this case, ties occur at three places. Students C and F have the same score and hence obtain the average rank of $(1+2)/2=1.5$. Students A and B have rank positions of 5 and 6 and hence are assigned $(5+6)/2=5.5$ each. Similarly students G and J have been assigned $(7+8)/2=7.5$ each. The same procedure is to be adopted when more than two persons tie up for the same position.

The rest of the procedure follows the same steps as have been used in the calculation of rho without ties. Difference of the ranks, D i.e. Col. (4) - Col. (5), has been calculated and summed to obtain $\sum D^2$, in Col. (6). This value has been inserted in the formula and the sum solved for the value of rho which is 0.855.

When ranks are treated as scores and there are no ties, the value of the product moment coefficient of correlation equals the value of rho. In case of tied positions, a correction may be used to make rho equal to r. However, in case of a small number of ties, the correction can be ignored and rho accepted as an approximation to r. Rank Difference Correlation is a quick and convenient method of estimating the correlation

TABLE 7.4

Rank-Difference Coefficient of Correlation

Student	Score on Test I <i>X</i>	Score on Test II <i>Y</i>	Rank on Test I <i>R</i> ₁	Rank on Test II <i>R</i> ₂	Differ- ence between Ranks <i>D</i>	Differ- ence squared <i>D</i> ²
A	10	16	6.5	5.5	1.0	1.00
B	15	16	3	5.5	-2.5	6.25
C	11	24	5	1.5	3.5	12.25
D	14	18	4	4	0	0
E	16	22	2	3	-1.0	1.00
F	20	24	1	1.5	-0.5	0.25
G	10	14	6.5	7.5	-1.0	1.00
H	8	10	9	10	-1.0	1.00
I	7	12	10	9	1.0	1.00
J	9	14	8	7.5	0.5	0.25
N=10					$\Sigma D^2=24.00$	

Formula:

$$\begin{aligned}
 \text{rho} &= 1 - \frac{6 \Sigma D^2}{N(N^2 - 1)} \\
 &= 1 - \frac{6 \times 24}{10(10^2 - 1)} \\
 &= 1 - \frac{144}{10 \times 99} \\
 &= 1 - \frac{144}{990} \\
 &= \frac{990 - 144}{990} \\
 &= 0.855
 \end{aligned}$$

Interpretation: The correlation between X and Y is very high and positive.

when N is small. In case only ranks are available, ρ is the only answer. With larger N 's, ρ may be useful as an exploratory measure. However, it must be amply evident to the student that r is based on both the sizes of the measures and also their relative positions in the series while ρ takes account of the positions only.

7.5 Properties of the Correlation Coefficient

7.5.1 The Range of r

The correlation coefficient may assume values from -1 through zero to $+1$. This is inherent in the very formula propounded by Pearson for the calculation of r . The values of $r = -1$ and $r = +1$ present a case of perfect relationship, though the direction of relationship is negative in the first case, and positive in the latter. The value of r can never be greater than $+1$ and less than -1 . These are the limits of r .

7.5.2 The Coefficient of Determination, r^2

The correlation coefficient, r , can be interpreted in terms of r^2 which is called the *coefficient of determination*. This may be called as the variance interpretation of r^2 . When multiplied by 100 the coefficient r^2 gives us the percentage of variance in Y that is associated with, determined by, or accounted for by variance in X . When $r = .60$, the value of $r^2 = (.60)^2 = .36$. Expressed in terms of percentage, it means that 36 per cent, $(.36 \times 100 = 36)$ of the variance in Y scores has been accounted for by the variance in X scores. The proportion of the variance in Y , not determined by or not associated with the variance in X is given by k^2 , which is called the *coefficient of non-determination*. Hence $k^2 = 1 - r^2$. Another index derived from the same is the *coefficient of alienation*, k .

$$k = \sqrt{1 - r^2} \quad (7.8)$$

while r indicates the degree of relationship between two variables, the coefficient of alienation, k indicates the degree of lack of relationship. Table 7.5 shows the three indices in relation to each other:

TABLE 7.5

Values of Coefficient of Determination, and Coefficient of Alienation for some Selected Values of r .

Correlation coefficient (r_{xy})	Coefficient of Determination, ($r^2_{xy} \times 100$)	Coefficient of Alienation (k)
.00	0.00	1.000
.05	0.00	1.000
.10	1.00	.999
.15	2.25	.989
.20	4.00	.980
.30	9.00	.954
.40	16.00	.917
.50	25.00	.866
.60	36.00	.800
.70	49.00	.714
.80	64.00	.600
.90	81.00	.436
.95	90.25	.312
.98	96.00	.199
.99	98.00	.141
.995	99.00	.100
.999	99.80	.045

An inspection of the above table reveals that the coefficients of determination for small r 's emphasize the very slight degree of association which these r 's disclose. An r of .10, .20 or even .30 between two tests X and Y indicates only 1 per cent, 4 per cent and 9 per cent, respectively, of the variance of Y accounted for by X. At the other extreme, r 's equal to .90 and .99 indicate 81 per cent and 98 per cent, respectively, of the variance in Y accounted for by X.

7.5.3 The Effect of Origin and Unit upon Correlation Coefficient

The value of r is invariant under transformations of unit and/or origin. *The correlation coefficient does not change if*

every score in either or both distributions is increased or multiplied by a constant. The result has important implications for the use of correlation coefficient. It leads us to conclude that it does not matter if the measurement is in feet or inches, minutes or seconds, units or dozens. The correlation between the variables will be the same. This quality of r gives it a large range of applications. While working with large values in the calculation of r by raw score method, it would always be advisable to subtract a constant from all the scores. It would help avoid working with large numbers.

7.5.4 Correlation and Causation

The correlation is sometimes misunderstood as indicating a causal relationship between the two variables and at times to the extent that the first was the cause, and the other, effect. However, such inferences are not legitimately possible. Although having a running nose correlates with having a cold, one would hardly believe that the running nose causes the cold. The fact the sun rises when we wake up in the morning does not suggest that our getting up causes the sun to rise.

If X correlates with Y , these causal relationships are possible:

- X causes Y ,
- Y causes X , or
- Z causes both X and Y .

Z may be very remote with long causal chain interposed before X and Y actually occur but the point serves the purpose of indicating a third source of causation. Hence, causality cannot be inferred solely on the basis of a correlation between two variables. It can be inferred only after conducting controlled experiments.

7.5.5. Factors Influencing the Size of the Correlation Coefficient

The student should be aware of the following factors which influence the size of the correlation coefficient and can lead to misinterpretation.

1. The size of r is very much dependent upon the variability of measured values in the correlated sample. The greater the variability, the higher will be the correlation, everything else being equal.
2. The size of r is altered when researchers select extreme groups of subjects in order to compare these groups with respect to certain behaviours. Selecting extreme groups on one variable increases the size of r over what would be obtained with more random sampling.
3. Combining two groups which differ in their mean values on one of the variables is not likely to faithfully represent the true situation as far as the correlation is concerned.
4. Addition of an extreme case (and conversely dropping of an extreme case) can lead to changes in the amount of correlation. Dropping of such a case leads to reduction in the correlation while the converse is also true.

7.5.6. Assumptions Underlying the Product Moment Correlation

Pearson's product moment r is based on some assumptions which must be fulfilled before its use is made. These assumptions include linearity of regression. It means that the trend of relationship between the two variables be rectilinear. This can be determined, as a rule, by inspection of the scatter diagram. If the distribution of the cases within the correlation diagram appears to be elliptical, without any indication of a definite bending of the ellipse, the chances are that the relationship is rectilinear. In case of curvature, of regression, *correlation ratio* instead of product moment r as a measure of relationship would be more appropriate. Curvilinearity of regression can be eliminated by transformations to binomial or to an approximately normal form.

Pearson's r does not assume that the distribution of the two variables should be normal. The forms of distributions may vary, so long as they are fairly symmetrical and unimodal. Even rectangular distributions can be used.

Many other circumstances affect the correlation coefficient. Among these may be mentioned sampling error and errors of measurement.

7.5.7. The Interpretation of r in Terms of Verbal Description

The correlation coefficient is generally interpreted in different ways by different statisticians. However, there is a fairly good agreement among them that the following verbal description may be assigned to different values of correlations.

<i>Value of r</i>	<i>Verbal description</i>
.00 to $\pm .20$	indifferent or negligible relationship
$\pm .20$ to $\pm .40$	low correlation; present but slight
$\pm .40$ to $\pm .70$	substantial or marked
$\pm .70$ to ± 1.00	high to very high

(Garrett, 1966)

The above categorization is broad and tentative and may be used with advantage as a general guide by less sophisticated students. However, a correlation coefficient must as a rule be judged with regard to

- (i) The nature of variables under study
- (ii) The statistical significance of the coefficient
- (iii) The degree of reliability of the tests used
- (iv) The purposes for which r has been computed, and
- (v) The extent of variability of the group.

7.6 Biserial Correlation

Some experiments require an estimate of the relationship between a continuous variable and a dichotomous variable. The term 'dichotomous' means 'cut into two parts.' The variable of social adjustment can be dichotomized as "socially adjusted subjects" and "socially maladjusted subjects." The two-fold classification may appear in the following example—Passes and failures; drop-outs and stay-ins; and successful and unsuccessful, etc. However, the essential assumption is that the variable underlying the dichotomy should be continuous and normal. It means that it should be based on artificial dichotomy and not on a natural dichotomy.

The formula for biserial correlation is

$$r_{bts} = \frac{M_p - M_q}{\sigma} \times \frac{pq}{y} \quad (7.9)$$

(calculation of r_{bts})

in which, r_{bts} = biserial r

M_p & M_q = mean test scores respectively for those who pass and fail the item

p & q = proportions who pass and fail the item

y = height of the ordinate of the normal curve at the point of division between p and q proportions of cases

σ = SD of the entire group

Example: A rehearsal group and a non-rehearsal group obtained the following scores on their performance. Find out the correlation between rehearsal and performance in dramatization.

TABLE 7.6
Worksheet for the calculation of r_{bts}

Scores	Rehearsal group	Non-rehearsal group
90—99	3	9
80—89	6	15
70—79	5	18
60—69	12	36
50—59	10	18
40—49	8	30
30—39	6	24
Sums	50	150

Steps:

1. Calculate Means for the two groups separately and combined. Also calculate SD for the total.

Group	Mean	SD
Rehearsal	60.9	(Mean and SD have been calculated using the standard procedures)
Non—rehearsal	59.5	
Total	59.85	
	17.63	

2. Obtain $p = N_1/N = 50/200 = .25$; $q = 1 - p = 1 - .25 = .75$
3. From Table M, ordinates of Normal curve, pick up the height of y ordinate corresponding to the point of dichotomy of .25 and .75. In our example, it is .318.
4. Substitute the values in Formula (7.9)

$$r_{bls} = \frac{60.9 - 59.5}{17.63} \times \frac{.25 \times .75}{.318}$$

$$= .047$$

The correlation is negligible.

An alternative formula for r_{bls} is

$$r_{bls} = \frac{M_p - M_q}{\sigma} \frac{p}{y} \quad (7.10)$$

(Alternative formula for r_{bls})

in which, M is the mean for the whole sample.

Values of r_{bls} may not always range between -1 and $+1$. In case of gross departure from normality, values of r_{bls} greater than unity may occur.

7.7 Point Biserial Correlation

In cases, when test items are scored simply as 1 if correct, and 0 if incorrect, that is right or wrong, the assumption of normality in the distribution of right wrong responses is generally not met. Other examples of genuine or natural dichotomies are, male-female, rural-urban, living-dead, convicted-not convicted, loyal-disloyal. In such cases, the point biserial correlation, r_{pbls} instead of r_{bls} can be used.

Example: Scores obtained by 11 students on the total test and item No. 12 of the test are given below in Table 7.7. Calculate item-total correlation.

Step 1

Calculate the following:

Proportion of students who passed the item, $p = \frac{6}{11} = .55$

Proportion of students who failed, $q = \frac{5}{11} = .45$

TABLE 7.7

Worksheet for the Calculation of Point Biserial
Correlation r_{pbis}

Score of test (Criterion) X	Item No. 12 Y	X^2
15	1	225
14	0	196
13	0	169
15	1	225
10	1	100
15	1	225
13	0	169
12	1	144
15	1	225
10	0	100
11	0	121
$\Sigma 143$	6	1899

M_x of students who passed the item,

$$M_p = \frac{15 + 15 + 10 + 15 + 12 + 15}{6} = 13.67$$

M_x of students who failed the item,

$$M_f = \frac{14 + 13 + 13 + 10 + 11}{5} = 12.20$$

$$M = \frac{\Sigma X}{N} = \frac{143}{11} = 13.00$$

Step 2

Calculate SD_x for the total group

$$= \frac{1}{N} \sqrt{\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{1}{11} \sqrt{11(1899) - (143)^2}$$

$$= 1.91$$

Step 3

Use the following formula to find out r_{pbis}

$$\begin{aligned} r_{pbis} &= \frac{M_p - M_q}{\sigma_y} \sqrt{pq} \\ &= \frac{13.67 - 12.20}{1.91} \sqrt{.55 \times .45} \\ &= .38 \end{aligned}$$

The point biserial correlation is a product-moment correlation. If we assign a 1 to individuals in one category and a 0 to individuals in the other category, and calculate product-moment r the result is a point biserial r . In the above example, those who did item No. 12 correctly were assigned a Y score of 1, and those who did it incorrectly a Y score of 0. Weights other than 1 and 0 can also be assigned. The point biserial r is specially useful in the analysis of the items of a test. The relation between biserial and point biserial correlation is given by the expression

$$r_{bis} = r_{pbis} \frac{\sqrt{pq}}{y}$$

The factor \sqrt{pq}/y varies from 1.25 when $p=q=.5$, to 3.71 when $p=.99$ and $q=.01$. Thus r_{bis} is always greater than r_{pbis} and the difference increases with extremeness of the dichotomies.

7.8 Tetrachoric Correlation

Point biserial and point biserial correlations were used when one variable was continuous and the other as test scores and the other as a dichotomy. A tetrachoric correlation is used when both variables are dichotomous. However, when both the variables are continuous, we can use the biserial or point biserial correlation. For example, if we wish to find out the correlation between two variables, we can use the point biserial correlation. For example, if we wish to find out

and punishment for indiscipline at school, we may dichotomize the variables as rural residence and urban residence; and punished and not-punished. Some other examples are:

1. Intelligence (Above average and below average); and social maturity (socially mature and socially immature);
2. School attendance (graduated from high school and those who did not); and employment (presently employed and not employed).

Tetrachoric correlation assumes that the two variables under study are essentially *continuous* and would be normally distributed if it were possible to obtain scores or exact measures on them.

Example: In the 2×2 table below the twofold distribution of students on training and success is given. Calculate tetrachoric correlation.

TABLE 7.8.

Worksheet for the calculation of tetrachoric correlation

Given	Pass	Fail	
Trained	20 (A)	40 (B)	60
Untrained	15 (C)	22 (D)	37
	35	62	97

$AD = 20 \times 22 = 440$

$BC = 40 \times 15 = 600$

Step 1: Both the variables are classified into two categories, marked + and -. Entries in cell A are + +, entries in D, --, so that concentration of frequencies into these two cells means close agreement and positive correlation. The other two cells are designated as B and C which would be + - and - +, or in our example Trained-Failed and Untrained-Passed. A generalized model of designating the cells is given below:

		X				X	
		+	-			-	+
Y	+	A ++	B -+	or Y	+	B -+	A ++
	-	C +-	D --		-	D --	C +-

Step 2: The full equation for tetrachoric r is *algebraically* very complex and hence a simplified formula which gives good approximation to r_t is used:

$$r_t = \cos \left(\frac{180^\circ \times \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}} \right) \quad (7.11)$$

(An approximate formula for tetrachoric r)

AD and BC are the products of the cells designated in the 2×2 table above and \cos is a trigonometric function whose value is available from tables.

In our example BC is greater than AD , hence AD is to be put in the numerator.

Substituting the values in Formula 7.11, we get,

$$\begin{aligned} r_t &= \cos \left(\frac{180^\circ \times \sqrt{440}}{\sqrt{440} + \sqrt{600}} \right) \\ &= \cos \left(\frac{3775.71}{45.47} \right) \\ &= \cos 83^\circ \end{aligned}$$

Step 3: Convert $\cos 83^\circ$ into r_t by consulting Table N in appendix.

$$r_t = -.122$$

(minus sign has been fixed to r_t because BC is greater than AD).

When the value of AD entries (agreement) is larger than the value of BC entries (disagreement), the correlation is

positive. In the reverse situation, the experimenter should fix a minus sign to the value of r_t .

7.9 The Phi Coefficient (ϕ)

When we have to find out correlation between two items of a test and the items are restricted to a scoring of 1 and 0, we can calculate phi coefficient instead of product-moment r . The data is arranged in a 2×2 table and the cells are marked as below:

Frequencies				Proportions				
		Item 2				Item 2		
		Fail	Pass			Fail	Pass	
Item 1	Pass	B	A	A + B	Item 1	(b)	(a)	p_1
	Fail	D	C	C + D		(d)	(c)	q_2
		B + D	A + C			q_1	p_1	

and then the following formulas are used:

$$\text{Phi coefficient} = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \quad (7.12)$$

(Phi coefficient from frequencies)

$$\text{Phi coefficient} = \frac{p_{ij} - p_i p_j}{\sqrt{p_i p_j q_i q_j}} \quad (7.13)$$

(Phi coefficient from proportions)

in which, p_i = proportion passing item i

p_j = proportion passing item j

q_i = proportion failing item i

q_j = proportion failing item j

p_{ij} = proportion passing both items i and j .

Example: The number of candidates passing and failing two items are given below. Calculate a Phi coefficient between the two items.

Solution

		Item I		
		Fail	Pass	
Item II	Pass	65 (B)	90 (A)	155 (A+B)
	Fail	80 (D)	30 (C)	110 (C+D)
		145 (B+D)	120 (A+C)	265

Substituting the values in Formula 7.12, we get

$$\begin{aligned} \text{Phi coefficient} &= \frac{90 \times 80 - 65 \times 30}{\sqrt{155 \times 110 \times 145 \times 120}} \\ &= \frac{7200 - 1950}{17224.11} = .30 \end{aligned}$$

The phi coefficient has been widely used in statistical work associated with psychological tests. Usually when researchers speak of the correlation between dichotomously scored test items, the reference is to the phi coefficient. The phi coefficient is a particular case of product moment correlation when the integers 1 and 0 are assigned to represent the two categories of each variable. The values of phi coefficient range between -1 and +1 and are influenced by the marginal totals. Negative and positive perfect correlation is obtained when the two variables are evenly divided; i.e., $p_i = q_i$ and $p_j = q_j$.

Exercises for Practice

- 7.1 What do you mean by 'Correlation'? Is it 'causation'?
- 7.2 What are the assumptions underlying product moment correlation? When should it be preferred to Rank Difference Correlation?
- 7.3 On what factors does the value of a correlation coefficient depend?

7.4 Calculate Pearson's Product Moment r from the following set of scores:

X	23	21	19	18	17	15	15	13	11	8
Y	15	19	13	17	15	13	12	11	10	5

7.5 Calculate Rank Difference Correlation from the following scores:

Maths.	55	58	51	53	48	49	52	59	60	54
Chemistry	61	47	39	38	36	43	49	50	42	41

7.6 Compute r on the following scores:

Intelligence	15	17	21	23	13	17	19	23	25	30
History	26	25	24	20	22	23	30	25	21	19

7.7. (a) If there are no ties in the scores what will happen if both r and ρ are calculated?

(b) If a constant of 5 is added to each of the scores of both the sets, what will be the effect on the value of r ?

(c) What are the uses of correlation?

7.8 Calculate Biserial r

Scores	Well adjusted f	Maladjusted f
45—49	0	6
40—44	3	5
35—39	4	5
30—34	6	10
25—29	2	8
20—24	3	6
15—19	1	10
10—14	1	10

7.9 Calculate Point Biserial r between total test and item No. 15

Student No.	Test Criterion(X)	Item No. 15
1	10	1
2	9	1
3	8	1
4	7	0
5	10	1
6	10	1
7	6	1
8	5	1
9	4	0
10	3	0
11	2	0
12	10	1

7.10 Calculate tetrachoric Correlation

		X Variable		
		Pass	Fail	
Y Variable	Trained	25	35	60
	Untrained	20	40	60
		45	75	120

7.11 Calculate Phi Coefficient

		Item No. I		
		Fail	Pass	
Item No. II	Pass	35	95	130
	Fail	60	70	130
		95	165	260

CHAPTER 8

THE SIGNIFICANCE OF MEAN AND OTHER STATISTICS

Inferential statistics is that branch of statistics which primarily deals with inferences from a sample to a larger population from which the sample has been taken. By implication, then, it is concerned also with the comparison of two sample estimates with a view to find out if they came from the same population or, in other words, if they did not differ significantly from each other on a given characteristic or property. A significant difference means a difference larger than expected by chance or due to sampling fluctuations. Means and other measures computed from samples are called *statistics* and are subject to chance difference due to sampling fluctuations. Measures descriptive of a population, on the other hand, are called *parameters* and are to be thought of as fixed reference values. We do not know the parameters, but they do exist. Under specified conditions, the parameters may be forecast from sample statistics with known degrees of accuracy. The degree to which a sample mean represents its parameter is an index of the *significance* or trustworthiness of the computed sample mean. When a statistic has been calculated, the question which is generally asked is: How good an estimate is this statistic of the parameter based upon the entire population from which this sample was drawn? This question applies to all statistics but only a few more important ones will be discussed in this chapter.

8.1 Sampling Distribution and the Standard Error of the Mean, SE_M

If a large number of samples are taken from the same

population and the same test administered to them under identical conditions, the average scores or means of these samples can be calculated. If the means so obtained are arranged in the form of a frequency distribution and also plotted on a graph as a frequency polygon, we obtain *the distribution of means*, which is called the sampling distribution of mean. The difference between a distribution of scores and a sampling distribution lies in the fact that the former is based on an arrangement of scores while the latter, of means or any other statistic. It has been found that even if the distribution of scores is skewed, the sampling distribution tends to reach a normal shape.

However, it may not be true in the case of very small samples. The smaller the sample, the more the form of distribution of the population affects the form of distribution of the means. It is important to have a knowledge of the form of sampling distribution of a statistic before we can draw any inferences from it about the parameters. It warrants the use of the theoretical models or theoretical mathematical distributions like binomial, normal, poisson and hypergeometric. However, in educational and psychological data, the normal distribution generally provides a good fit and hence is most popularly used. *The Standard Error (SE) is the standard deviation of the sampling distribution* and is to be interpreted in the same manner. The sampling distributions, though, are not calculated, yet they exist. The SE is also not calculated direct from the sampling distribution, but it is estimated from the sample standard deviation which is the only value available to us.

8.2 Computation of the Standard Error of the Mean, SE_M

For the computation of SE_M , we need to know the population standard deviation and then to use the following formula:

$$\sigma_M = \frac{\sigma}{\sqrt{N}} \quad (8.1)$$

(SE of a Mean computed from a known population parameter)
in which σ = SD of the population; N = number of cases in the sample.

However, the population parameter, σ , is generally unknown and cannot be directly obtained experimentally. It may involve a huge expense in terms of money and time to test the whole population and may defeat the very purpose of the experiment itself. Hence statisticians have devised methods of estimating σ_M from the sample statistics available to the experimenter. The formula for the purpose is

$$\text{SE of the Mean, } SE_M, \sigma_M = \frac{\sigma}{\sqrt{N}} \quad (8.2)$$

(SE of the Mean estimated from sample standard deviation) in which, σ is the sample SD; and N is the number of cases in the sample.

Some authors suggest the use of N in the denominator of Formula (8.2) for large samples ($N \geq 30$) and of $N-1$ for small samples ($N < 30$). The plea is, that in very large samples generally used in social sciences, no appreciable difference takes place in the value of σ_M by $N-1$ instead of N . The use of N or $N-1$ thus remains a matter of arbitrary decision. However if $N-1$ instead of N has not been used in calculating the sample SD, then it becomes imperative to use $N-1$ instead of N in the denominator of the formula (8.2) for obtaining an unbiased estimate of the population standard deviation, σ . It has been shown that SD of a random sample underestimates (is smaller than) the corresponding population σ . Hence for the correction of this underestimation, the SD of a sample should be computed by the formula

$$SD = \sqrt{\frac{\sum x^2}{N-1}} \text{ instead of the usual formula } \sqrt{\frac{\sum x^2}{N}}$$

The student will easily understand that sample σ is the only estimate of the σ available to us and hence the former be used in the calculation of the SE of the mean.

Formula (8.2) above makes it clear that the size of the SE_M varies directly with the size of the sample SD and inversely with the size of N .

8.3 Application and Interpretation of SE_M In Large Samples

Standard Error of mean measures the degree to which the mean is affected by the errors of measurement as well as by the errors of sampling or sampling fluctuations from one random sample to the other. The interpretation of the SE of the mean is done to answer the question — how dependable is the mean? Further it may be asked, as to how good an estimate is the sample mean of the population mean. The answer to these questions is provided by setting up the *confidence limits* or the *fiduciary limits* of the mean. These limits, for a particular level of confidence, are supposed to embrace the population mean. Thus the interpretation of the SE_M is done in terms of the *confidence intervals* for the population mean.

Example: The mean achievement score of a random sample of 400 students on a test of statistics is 57 and SD is 15. How dependable is the mean? How good an estimate is it of the population mean?

Solution

Step 1. Calculate SE_M by Formula (8.2)

$$SE_M = \frac{\sigma}{\sqrt{N}} = \frac{15}{\sqrt{400}} = \frac{15}{20} = .75$$

We have used SD of the sample as our estimate of the population σ . SE_M is the standard deviation of a distribution of sample means around the fixed population mean (population mean is a constant value).

Assuming normal distribution the position of the SE_M is shown in Figure 8.1.

Step 2. Set up Confidence Intervals

In Figure 8.1, the population mean which is unknown is at the centre of the curve. Normal Curve, being symmetrical, sample means fall equally often on the + (upper) side and — (lower) side of the M_{Pop} . Looking at the divisions of area, 68.26 per cent or about $2/3$ of the sample means fall between

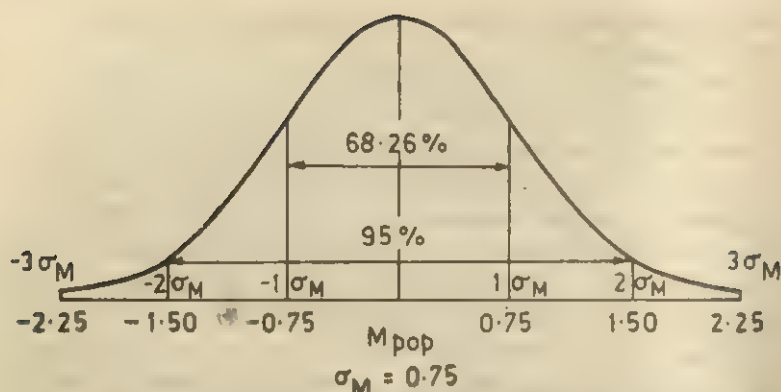


Fig. 8.1 Sampling Distribution of Means, showing Variability of Obtained Means around Population M in terms of σ_M .

$\pm 1\sigma_M$ of M_{pop} i.e. within a range of $\pm 1 \times .75$ units. Proceeding further out towards the two tails of the curve, one may notice that 95 per cent sample means lie within $\pm 2\sigma_M$ (More exactly $1.96\sigma_M$; See Table A in appendix).

For setting up fiduciary* limits or confidence intervals one may proceed as follows.

At .05 level, value of $z=1.96$ (from Normal Curve Table A).

$$\begin{aligned}
 & M \pm 1.96\sigma_M \\
 & = 57 \pm 1.96 \times .75 \\
 & = 57 \pm 1.47 \\
 & = 55.53 \text{ to } 58.47
 \end{aligned}
 \tag{8.3}$$

At .01 level, value of $z=2.58$ (from Normal Curve Table A)

$$\begin{aligned}
 & M \pm 2.58\sigma_M \\
 & = 57 \pm 2.58 \times .75 \\
 & = 57 \pm 1.94 \\
 & = 55.06 \text{ to } 58.94
 \end{aligned}
 \tag{8.4}$$

*R.A. Fisher termed the confidence intervals of a parameter as *fiduciary limits* and the confidence placed in the interval defined by the limits as containing the parameter, as *fiduciary probability*.

The above confidence intervals i.e at .05 and at .01 levels, are in general use accepted as standard by most of the statisticians. However, confidence intervals with lesser degrees of assurance can also be set up.

Step 3. Interpret the Results

The confidence intervals represent a range within which the parameter, M_{Pop} is likely to fall. Hence, with respect to our data above, there are 95 chances out of 100, that the M_{Pop} would fall between the score values 55.53 to 58.47; and there are 99 chances out of 100, that the M_{Pop} would fall between the score limits 55.06 to 58.94. Our confidence that these intervals contain M_{Pop} is 95 per cent or P of .95; and 99 per cent or P of .99 respectively.

8.4 The Distribution of t

The t distribution is a theoretical distribution discovered by an English, statistician, W.S. Gossett in 1908 writing under the pen name 'Student' in respect of his teacher R.A. Fisher. The distribution is therefore known as 'Student's' distribution. The t ratio is obtained by the formula

$$t = \frac{M - \mu}{S_M} \quad (8.5)$$

(Basic Formula of t ratio)

in which μ is the population mean; M, mean of the sample; and S_M , an estimate of the σ_M obtained from SD of the sample. Gossett had shown that for random samples drawn from normal populations, the sampling distribution of t is given by

$$y = \frac{y_0}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} \quad (8.6)$$

(Equation for sampling distribution of t)

in which n is the number of cases in the sample and y stands for length of the ordinate. The denominator of the equation will be minimum if $t=0$, and hence the height of the ordinate

the maximum at this point. Since the squared value of t is used, the ordinate y will be the same for positive and negative values and will generate a symmetrical distribution. Finally, as t increases, the ordinate y , or the height of the curve decreases. The curve is asymptotic to the base line. The curve is much like the normal curve except that it is more peaked for small n 's. As the n 's grow in size, the t distribution approaches normality.

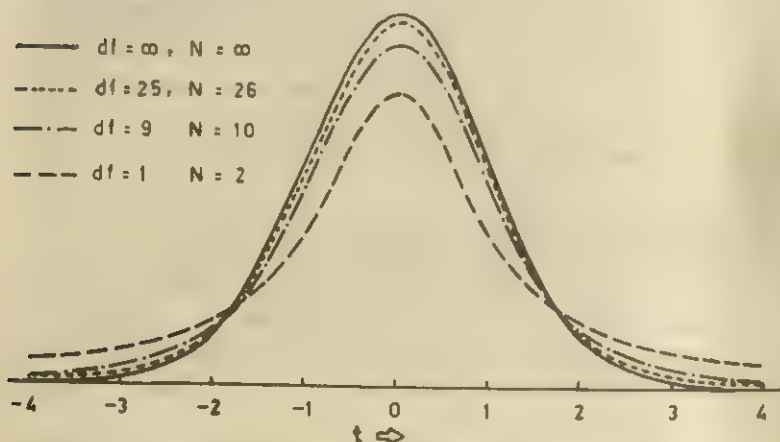


Fig. 8.2. Distribution of t for different degrees of freedom ranging from 1 to ∞ .

As depicted in Figure 8.2, the t distribution is not a single distribution but a family of distributions depending on df . Like binomial and normal distributions already discussed in a previous chapter, t distribution is another theoretical model having wide applications to several sampling problems. Tables of values of t have been devised by statisticians. Table B in appendix shows these values for different levels of confidence and is very simple to use. The value of t at the intersection of a given df and level of confidence is to be picked up as the critical value.

8.5 Degrees of Freedom, df

The concept of degrees of freedom is of key importance in inferential statistics. Almost all tests of significance require

the calculation of degrees of freedom. It is a mathematical concept. The geometric interpretation of the concept relates to the movement of a point in relation to the number of dimensions it is attached with. A point on a line is free to move in one dimension only and thus has 1 degree of freedom. A point on a plane has freedom of movement in two dimensions, and has 2 degrees of freedom. A point in space of three dimensions has 3 df.

The expression degrees of freedom is abbreviated from the full expression "degrees of freedom to vary". When a sample statistic is used to estimate a parameter, the number of degrees of freedom depends upon the number of restrictions placed upon the scores, each restriction reducing one df. To a less mathematically oriented student, the idea of df can be presented through two sets of scores given below

	Set I	Set II	
<i>Sr. No.</i>	<i>Given Scores</i>	<i>Altered Scores</i>	
1.	8	10	Can vary in any way.
2.	7	6	
3.	6	12	
4.	10	13	
5.	19	9	— Its value gets fixed or automatically determined.
ΣX	50	50	
M_x	10	10	

In set I original scores have been presented. We can vary the first four scores, 1-4, in any way we like. But the value of the 5th score gets fixed due to the restriction that ΣX in each case must be 50. Hence in this situation the four scores out of five are free to vary and hence the $df=5-1=4$. Thus in this case, the number of df depended upon the number of scores minus the number of restrictions (The idea of df has also been explained in chapter on chi-square).

8.6 Levels of Significance

Experimenters and researchers have selected some arbitrary standards—called *levels of significance* to serve as the *cut-off points or critical points along the probability scale, to separate the significant difference from the non-significant difference* between the two statistics, like means or SD's. Generally, the .05 and the .01 levels of significance are the most popular in social sciences research. The confidence with which an experimenter rejects—or retains—a null hypothesis depends upon the level of significance adopted. These may, hence, sometime be termed as *levels of confidence*. Their meanings may be clear from the following:

TABLE 8.1

Meanings of Levels of Confidence

<i>Level</i>	<i>Amount of confidence</i>	<i>Interpretation</i>
.05	95%	If the experiment is repeated a 100 times, only on five occasions the obtained mean will fall outside the limits $\mu \pm 1.96 \text{ SE}$
.01	99%	If the experiment is repeated a 100 times, only on 1 occasion, the obtained mean will fall outside the limits $\mu \pm 2.58 \text{ SE}$

The values 1.96 and 2.58 have been taken from the *t* tables keeping large samples in view. The .01 level is more rigorous and higher a standard as compared to the .05 level and would require a larger value of the critical ratio for the rejection of the H_0 . Hence if an obtained value of *t* is significant at .01 level, it is automatically significant at .05 level but the reverse is not always true.

8.7 Application and Interpretation of SE_M in Small Samples

The procedure of calculation and interpretation of Standard Error of Mean in small samples differs from that for large samples, in two respects.

1. The denominator $N-1$ instead of N is used in the formula for calculation of the SD of the sample.
2. The appropriate distribution to be used for small samples is t distribution instead of normal distribution.

The rest of the line of reasoning used in determining and interpreting SE_M in small samples is similar to that for the large samples.

Example: A randomly selected group of 16 students was administered a test of verbal ability. The mean and SD obtained by the group are 52 and 8. Determine the 95 per cent and 99 per cent confidence intervals for the M_{Pop} .

Solution

Step I Calculate SE_M

$$SE_M = \frac{\sigma}{\sqrt{N}} = \frac{8}{\sqrt{16}} = 2.00$$

Step II From Table B pick up the values of t

for $df = N - 1 = 16 - 1 = 15$:

t at .05 level = 2.13

and t at .01 level = 2.95

Step III Set up confidence intervals

at .05; $M \pm 2.13 SE_M$

$$= 52 \pm 2.13 \times 2.00$$

$$= 47.74 - 56.26$$

at .01; $M \pm 2.95 SE_M$

$$= 52 \pm 2.95 \times 2.00$$

$$= 46.10 - 57.90$$

Step IV Interpret the results

There is a probability of .95 that the M_{Pop} will be within the score range 47.74 to 56.26; and a .99 probability that M_{Pop} will be within the score range 46.10–57.90.

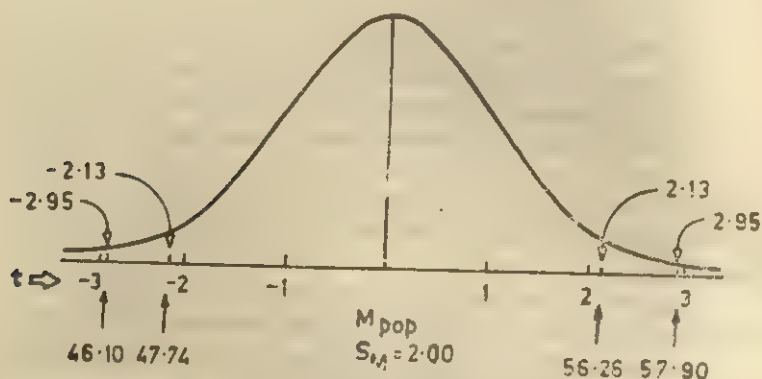


Fig. 8.3. Confidence Intervals for the M_{Pop} in the t Distribution with $df=15$.

A look at Figure 8.3 will reveal the placement of the limits of the two confidence intervals in terms of SE_M and scores. The width of the .99 confidence interval (i.e. 46.10–57.90) is larger than the .95 confidence interval (i.e. 47.74–56.26). If the experiment or observation is repeated over a large number of times, chances are that the M_{Pop} will be within 46.10–57.90 with our chances of being correct 99 per cent times and being wrong only once in 100 times. Small n 's do not necessarily generate stability of the results because smallness of the sample may not lead to an accurate representation of the parent population. Randomness in such cases may prove ineffective to guarantee this condition.

8.8 The Standard Error of a Median, σ_{Mdn}

It has been established that the variability of the sample medians is about 25 per cent greater than the variability of means in a normally distributed population. Hence the standard error of a median can be estimated by using the formulas:

$$SE_{Mdn}, \sigma_{Mdn} = \frac{1.253\sigma}{\sqrt{N}} \quad (8.7)$$

$$\sigma_{Mdn} = \frac{1.858Q}{\sqrt{N}} \quad (8.8)$$

(Standard Error of the Median in terms of σ and Q)

It is clear from Formula (8.7) above that SE_{Mdn} is roughly 4 times the SE_M . Hence σ_{Mdn} is less dependable and more subject to sampling fluctuations as compared to σ_M .

Example: On a test of mechanical aptitude 225 randomly selected students of an engineering course secured a $Mdn = 25.50$ and $Q = 5.00$. How well does this median represent the median of the population from which this sample was drawn?

Solution

By Formula 8.8, the $\sigma_{Mdn} = \frac{1.858 \times 5.00}{\sqrt{225}} = .62$

Since N is large, we can use normal curve tables to set up the confidence intervals.

z for 99 per cent = 2.58

The same values we used in

z for 95 per cent = 1.96

large sample SE_M cases.

Hence the confidence intervals are:

$$\begin{aligned} .99 \text{ confidence: } & Mdn \pm 2.58 \sigma_{Mdn} \\ & = 25.50 \pm 2.58 \times .62 \\ & = 25.50 \pm 1.60 \\ & = 23.90 \text{ to } 27.10 \end{aligned}$$

$$\begin{aligned} .95 \text{ confidence: } & Mdn \pm 1.96 \sigma_{Mdn} \\ & = 25.50 \pm 1.96 \times .62 \\ & = 25.50 \pm 1.22 \\ & = 24.38 \text{ to } 26.72 \end{aligned}$$

The interpretation of the confidence intervals follow the same pattern as that of the mean. Here we can place 99 per cent and 95 per cent confidence respectively that the Population mdn. will be within these ranges.

8.9 The Standard Error of a Standard Deviation, SE_{σ}

Since standard deviation also records fluctuations from sample to sample, the SE of the standard deviation can also be estimated and used to find the limits within which population SD will fall. The formula for the purpose is

$$SE_{\sigma}, \sigma_{\sigma} = \frac{\sigma}{\sqrt{2N}} \quad (8.9)$$

(Standard Error of a Standard Deviation)

For small samples with $N < 100$, the sampling distribution of SD is somewhat skewed but approaches normality as N increases. Hence inferences based on normal distribution can be drawn. A comparison of the denominators of Formula (8.9) and (8.2) will reveal that the SE_M is about 40 per cent greater than SE_{σ} and hence less stable than the SE_{σ} .

After calculating SE_{σ} we can set up the confidence intervals as explained earlier.

8.10 The Standard Error of Percentages and Proportions

At times it is not possible to measure some traits. The only information available is the percentage of the group that possess that trait. The SE of the percentage would then be required to estimate the degree of confidence we can place in our information. How reliable was the percentage as an index of the incidence of the behaviour in which we are interested? The formula for the SE of the percentage is

$$SE_{\%}, \sigma_{\%} = \sqrt{\frac{PQ}{N}} \quad (8.10)$$

(Standard Error of Percentage)

in which P = percentage of the group possessing the trait

$Q = (1 - P)$

N = Number of cases.

Suppose, 80 or 40 per cent of the 200 children were found to have shown signs of tiredness after a strenuous physical exercise. Assuming that the sample was randomly drawn from a specified population, how well do our results represent the population percentage.

Applying Formula (8.10) we get

$$SE_{\%} = \sqrt{\frac{40\% \times 60\%}{200}} = 3.46\%$$

The .99 confidence interval will be : $40\% \pm 2.58 \times 3.46\%$
 $= 40\% \pm 8.93\%$

$$= 31.07\% \text{ to } 48.93\%$$

The .95 confidence interval will be $40\% \pm 1.96 \times 3.46\%$

$$= 40\% \pm 6.78\%$$

$$= 33.22\% \text{ to } 46.78\%$$

We may feel sure with 99 per cent confidence that the percentage of children in the population who are likely to be tired after the physical exercise will not be less than 31 per cent and more than 48.93 per cent.

8.11 The Standard Error of a Correlation Coefficient, SE_r

We may draw a large number of samples randomly from a population, compute a correlation coefficient for each sample, and prepare a frequency distribution of correlation coefficients. The shape of this distribution depends upon r_{Pop} . As r_{Pop} departs from zero, the sampling distribution of r 's becomes increasingly skewed. A high positive value of r_{Pop} generates an extremely negative skewness, while a high negative value of r_{Pop} produces an extremely positive skewness. The SE of r is given by the formula

$$SE_r, \sigma_r = \frac{1-r^2}{\sqrt{N}} \quad (8.11)$$

(SE of a correlation coefficient)

For example, the value of r in a set of 100 scores is .6. How dependable is this value?

$$\text{By Formula (8.11), we have, } \sigma_r = \frac{1-.6^2}{\sqrt{100}} = \frac{.64}{10} = .064$$

Using normal distribution, we set up the confidence intervals as below:

$$\begin{aligned} .99 \text{ confidence } & r \pm 2.58 \sigma_r \\ & = .6 \pm 2.58 \times .064 \\ & = .6 \pm .165 \\ & = .435 - .765 \end{aligned}$$

$$\begin{aligned} .95 \text{ confidence } & r \pm 1.96 \sigma_r \\ & = .6 \pm 1.96 \times .064 \\ & = .6 \pm .125 \\ & = .475 - .725 \end{aligned}$$

The interpretation is that the Pop. r will be within these limits.

This formula has a fundamental defect that the theoretical model of normal distribution is not a good fit to the sampling distribution of r 's as explained earlier. Hence z transformation of r is recommended.

8.12 Conversion of r 's Into Fisher's z Function

Difficulties resulting from the non-normality of the sampling distribution of r were resolved by R.A. Fisher who suggested the conversion of r 's into z function by using the formula.

$$z_r = \frac{1}{2} \log_e (1+r) - \frac{1}{2} \log_e (1-r) \quad (8.12)$$

(conversion of r 's into Fisher's z function)

However, conversion tables are readily available and there is no need of using Formula (8.12). See Table C in appendix. The given r 's are converted into z 's. The test of significance is then applied to z and not to r 's. The sampling distribution of z is approximately normal and the values of z 's can be interpreted in this manner. The SE formula for z , as given below is independent of the value of r 's.

$$SE_z = \frac{1}{\sqrt{N-3}} \quad (8.13)$$

(SE for Fisher's z Function)

For illustration, the problem mentioned above is taken in which $r=.6$ and $N=100$.

1. Convert r of .6 into z which $=.69$
2. By Formula 8.13, $SE_z = \frac{1}{\sqrt{N-3}} = \frac{1}{9.85} = .1$
3. Set up the confidence intervals as below:

For .99 confidence: $z \pm 2.58 \sigma_z$
 $=.69 \pm 2.58 \times .1$
 $=.69 \pm .26$
 $=.43 \text{ to } .95$

Convert back into r 's $=.405 \text{ to } .740$

For .95 confidence: $z \pm 1.96 \sigma_z$
 $=.69 \pm 1.96 \times .1$
 $=.69 \pm .196$
 $=.494 \text{ to } .786$
 $=.49 \text{ to } .79 \text{ (rounded)}$

Convert back into r 's $=.455 \text{ to } .660$

A comparison of these values with those obtained by Formula (8.11) would reveal some differences but not very appreciable.

In addition of the above, SE formulas for some other statistics and in some special situations have also been suggested. Only formula for SE of the quartile deviation is given:

$$SE_Q, \sigma_Q = \frac{.786 \sigma}{\sqrt{N}} \quad (8.14)$$

Based on Q

$$SE_Q, \sigma_Q = \frac{1.17 Q}{\sqrt{N}} \quad (8.15)$$

Exercises for Practice

- 8.1 On a test of clerical aptitude, a randomly selected group of 400 candidates obtains a mean $=25.6$ and $\sigma=5.00$. Use .95 and .99 confidence, and set up the two confidence intervals.

- 8.2 A randomly selected group of 25, VI grade students have a mean height of 135 cm, and $SD=10$ cm. How well does this value estimate the population mean? Use .99 and .95 confidence intervals.
- 8.3 The mean of a large randomly selected sample is K and $\sigma_K=3.00$. What are the chances that the sample mean misses the true mean by (a) ± 1.00 ; (b) ± 3.00 ; (c) ± 10.00 .
- 8.4 In a sample of 625 voters, 55 per cent favour a particular political party A. How dependable is this percentage?
- 8.5 The mdn. score of 225 students randomly selected is 24.5 and $Q=4.2$. Set up the fiduciary limits within which the population mdn. is likely to fall. Use .99 and .95 confidence levels.
- 8.6 The standard deviation of the intelligence scores of a group of 200 randomly selected students is 10.2. How dependable is the SD?
- 8.7 An r of .74 is obtained from a random sample of 39 cases. Use z conversion and set up the fiduciary limits of the .99 and .95 confidence intervals.
- 8.8 (a) What do you mean by the dependability of a statistic?
(b) What are fiduciary limits?
(c) What is the purpose of inferential statistics?
- 8.9 Compare normal and t distributions. When do they coincide?
- 8.10 Why is a theoretical distribution model necessary for estimation?
- 8.11 Explain the concepts of degrees of freedom and confidence levels.

CHAPTER 9

THE SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS AND OTHER STATISTICS

In experimental and other research work, the determination of whether an observed difference is of such magnitude that it cannot be attributed to chance factors or sampling variations, is often our major interest. For example, we may observe that a group of subjects tested under one set of experimental conditions has a higher mean than a comparable group tested under a different set of experimental conditions. Is the observed difference between the means one that might occur frequently by chance, that is, as a result of sampling variations? If not, then we might infer that the difference is a product of the experimental conditions. For this purpose, we need a statistical test of significance of difference between the means. The critical ratio or the t test is the one generally used in such circumstances.

9.1 The Null Hypothesis, H_0

The *null hypothesis* is a proposition of zero differences. Fisher has emphasized that every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis. Thus a hypothesis which is set up with the possibility of its being rejected at some defined probability value is called a null hypothesis, the term "null" referring to our interest in the possible rejection of the hypothesis. In statistical terms a null hypothesis may be stated as

$$H_0: \mu_1 = \mu_2,$$

where μ_1 and μ_2 are population means.

It states that there is no significant difference in the means of the two populations. If facts lead to the rejection or non-retention of the H_0 , then the alternative hypothesis (H_1) as stated below, stands accepted.

$H_1: \mu_1 \neq \mu_2$; in which \neq means "not equal to"; and μ_1 and μ_2 are population means.

It means that the two populations differ significantly.

9.2 The Process

The process of testing for the significance of the difference between the two means includes the following steps:

- (i) Set up H_0 and the H_1 according to the requirements of the problem.
- (ii) Decide about the level of significance for the test. Customarily .05 and .01 levels are selected.
- (iii) Decide whether one-tailed or two-tailed test of significance was needed.
- (iv) Decide whether the data warranted a test of significance for the independent or the correlated means.
- (v) Decide whether the large sample or the small sample was involved.
- (vi) Use one of the following formulas appropriate to (iv) and (v) above, for the calculation of SE of Mean Difference, SE_D . (See page 185)
- (vii) Calculate the value of the critical ratio or t by using the formula $\frac{M_1 - M_2}{SE_D}$ in which M_1 & M_2 are the means to be compared and SE_D is the SE of Mean Difference calculated under step VI above.
- (viii) Calculate Degrees of Freedom (df) as below:
 - (a) For Uncorrelated or Independent Samples, $df = N_1 + N_2 - 2$.
 - (b) For Correlated Samples, $df = N - 1$.
- (ix) Look up the tables of t values (See Appendix, Table B) with df (as decided in step VIII above) and the level of significance (step II).

1. Independent or Uncorrelated Means Large Samples

$$SE_D, \sigma_D = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2}}$$

or

$$\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \quad (9.1)$$

Small Samples

$$SE_D = SD \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \quad (9.2)$$

$$\text{Where } SD = \sqrt{\frac{\sum (X_1 - M_1)^2 + \sum (X_2 - M_2)^2}{(N_1 - 1) + (N_2 - 1)}}$$

2. Correlated Means Large Samples

$$SE_D = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2} - 2r_{12} \sigma_{M_1} \sigma_{M_2}} \quad (9.3)$$

Small Samples

$$SE_D = \frac{SD}{\sqrt{N}} \quad (9.4)$$

Description of symbols

σ_{M_1} and σ_{M_2} = SE's of the Means of the two groups.

σ_1 and σ_2 = SD's of the two groups.
 N_1 and N_2 = Number of cases in the two groups.

SD = Pooled Standard Deviation of the two groups.

Other symbols as above.

M_1 & M_2 = Means of the two groups.
 X_1 & X_2 = Individual raw scores in the two groups.

r_{12} = Correlation Coefficient between scores of Groups I and II.

Other symbols as above.

SD = The standard deviation of the Difference Scores.

N = No. of Differences or
No. of persons in the group.

- (x) **Decision Rule :** Compare the calculated value of t .
 - (a) If the calculated value of t is larger than the table value of t , reject H_0 and accept H_1 .
 - (b) If the calculated value of t is less than the table value of t , accept H_0 .
- (xi) **Interpret the results as below:**
 - (a) H_0 rejected: There is a significant difference between the two means.
 - (b) H_0 accepted: There is no significant difference between the two means. Whatever the difference, it has arisen due to sampling fluctuations and chance factors only.

The procedural steps in the use of the t test of significance of difference between two means will now be explained with the help of some numerical examples. For the purpose of convenience, the examples have been arranged in two sections. The first section is concerned with the comparison of independent or uncorrelated means. The second section is on comparison of correlated means.

9.3 Standard Error (SE) of the Difference between two Independent Means (Large Sample)

When two distinct groups of subjects are involved, the groups may be termed as independent. These groups are drawn at random from totally different and unrelated populations. No attempt is made to equate the groups by using pair-comparison or any other method.

Example: Thirty boys and forty girls selected randomly from the eighth class of a big school were given a standard test of Arithmetic Ability. Their means and SD's are reported below:

	N	M	SD
Boys	30	20.5	4.0
Girls	40	16.2	5.0

Is the mean difference in the arithmetic ability significant?

TABLE 9.1

**Summary of the Test of Difference of Means of
Independent Groups (Arithmetic Ability Example)**

Hypothesis

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Decision Rules

Given : .05 significance level; & $df = N_1 + N_2 - 2 = 68$;
table value of $t = 2.00$.

If $t_{\text{obt.}} < 2.00$ accept H_0 .

If $t_{\text{obt.}} \geq 2.00$ reject H_0 .

Computation

$$\text{Formula, } SE_D = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Substituting the numerical values.

$$SE_D = \sqrt{\frac{(4.0)^2}{30} + \frac{(5.0)^2}{40}} = 1.07$$

$$t = \frac{M_1 - M_2}{SE_D}$$

$$\frac{20.5 - 16.2}{1.07} = \frac{4.3}{1.07} = 4.02$$

Interpretation : Reject H_0 .

9.4 The SE of Difference between Means in small Independent Samples

Example : An attitude test was administered to 10 boys in English class and 5 boys in Hindi class. Their scores are given below. Is the mean difference between the groups significant?

TABLE 9.2

**Summary of the test of Difference between Means
in small Independent Samples
(Attitude test example)**

Hypothesis: $H_0 : \mu_1 = \mu_2$; $H_1 : \mu_1 \neq \mu_2$

Decision Rules:

Given significance level = .05 & $df = N_1 + N_2 - 2 = 13$,
and table value of $t = 2.16$

If calculated value of $t < 2.16$, accept H_0

If calculated value of $t \geq 2.16$ reject H_0

Computation

English Course			Hindi Course		
X	x	x^2	X	x	x^2
6	-4	16	4	1	1
7	-3	9	3	0	0
8	-2	4	2	-1	1
10	0	0	1	-2	4
15	+5	25	5	2	4
16	+6	36			
9	-1	1			
10	0	0	$\Sigma X_2 = 15$		$\Sigma x_2^2 = 10$
10	0	0			
9	-1	1	$M = 15/5 = 3$		
			$N_1 - 1 = 9, N_2 - 1 = 4$		
$\Sigma X_1 = 100$		$\Sigma x_1^2 = 92$			
$M = 10$					

$$\text{Formula: } SD = \sqrt{\frac{\Sigma x_1^2 + \Sigma x_2^2}{(N_1 - 1) + (N_2 - 1)}} = \sqrt{\frac{92 + 10}{9 + 4}} = 2.8$$

$$SE_D = SD \sqrt{\frac{N_1 + N_2}{N_1 N_2}} = 2.8; \sqrt{\frac{10 + 5}{10 \times 5}} = 1.54$$

$$t = \frac{M_1 - M_2}{SE_D} = \frac{10 - 3}{1.54} = 4.46$$

Interpretation:

Reject H_0 : It shows that the two groups differed significantly on their mean attitude scores.

9.5 Standard Error of the Difference between Two Correlated Means

Correlation between the two means is introduced in the following situations:

- (a) *The Single Group Situation:* When a single group is tested twice on the same test or an equivalent form of the test is used on the second occasion.
- (b) *The Equivalent Group Situation:* When equivalent groups are formed by using "matching by pairs" or "matching by Means and SD's"

If the same group of students takes the arithmetic ability test twice instead of two different groups taking it, we have the same individual's score on the first testing to pair off with his score in the second testing. If in a comparison of males and females, the two groups are standardized better by taking a brother or a sister from each family or if the boys and girls are paired with respect to age, IQ, or social status and if these factors of common family, age, IQ or social status have any relation to arithmetic ability, they will automatically introduce correlation between the two samples.

A correlation coefficient is computed and introduced in the relevant formula. A numerical example using the single group situation is given below:

TABLE 9.3

Summary of the Test of the Difference between Means for Correlated Large Samples (Single Group Method)

Example: A group of 35 randomly selected students was tested before and after an experimental treatment. The data so obtained is given below:

	<i>Pre-test</i>	<i>Post-test</i>	
M	15.5	21.6	$r=.70$
SD	5.2	4.8	$N=35$

Find out if the groups differed significantly on the two testings.

Hypothesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Decision Rules: Given significance level = .01 & $df = N - 1 = 34$; and table value of $t = 2.72$

If calculated $t < 2.72$, accept H_0

If calculated $t \geq 2.72$, reject H_0

Computation:

$$\text{Formula } SE_D = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2 - 2 r_{12} \sigma_{M_1} \sigma_{M_2}}$$

$$\sigma_{M_1} = \frac{\sigma_1}{\sqrt{N}} = \frac{5.2}{\sqrt{35}} = .88 \quad (\sigma_1 \text{ is the SD of pre-test})$$

$$\sigma_{M_2} = \frac{\sigma_2}{\sqrt{N}} = \frac{4.8}{\sqrt{35}} = .81 \quad (\sigma_2 \text{ is the SD of post-test})$$

$$SE_D = (.88)^2 + (.81)^2 - 2 \times .70 \times .88 \times .81 = .658$$

$$t = \frac{M_1 - M_2}{SE_D} = \frac{6.1}{.658} = 9.27$$

Interpretation:

Reject H_0

There is a significant difference between the mean scores of the group on pre-test and post-test.

9.6 Difference Method (small samples)

When groups are small, the procedure called, the *difference method* is often to be preferred. It is quicker and easier to apply than the long method of calculating SE's for each mean and the SE of the difference. It is to be preferred if the value of the correlation coefficient between the two sets of scores is not required for any other purpose.

Example: Ten subjects were given three successive trials on a non-sense syllable test. The scores for the first and the last trials are shown below. Is the mean gain from the first to the third trial significant?

TABLE 9.4

**Summary of the Test of the Difference between Means
for Correlated Groups (Non-sense syllable test
example) Difference Method**

Hypothesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

*Decision Rules*Given significance level = .01; & $df = N - 1 = 9$ and table value of $t = 3.25$ If calculated $t < 3.25$ accept H_0 If calculated $t \geq 3.25$, reject H_0 *Computation:*

<i>Trial-I</i>	<i>Trial-III</i>	<i>Difference (T-III—T-I)</i>	<i>x</i>	<i>x²</i>
12	16	4	1	1
14	18	4	1	1
10	17	7	4	16
8	10	2	-1	1
16	18	2	-1	1
17	25	8	5	25
18	18	0	-3	9
20	21	1	-2	4
16	17	1	-2	4
19	20	1	-2	4
<hr/> Σ 150	<hr/> 180	<hr/> ΣD 30	<hr/>	<hr/> Σx^2 66

$$\text{Mean } D = \frac{\Sigma D}{N} = \frac{30}{10} = 3.0$$

$$SD_D = \sqrt{\frac{\Sigma x^2}{N-1}} = \sqrt{\frac{66}{9}} = 2.71$$

$$SE_D = \frac{SD}{\sqrt{N}} = \frac{2.71}{\sqrt{10}} = .86$$

$$t = \frac{M_D - 0}{SE_D} = \frac{3.0}{.86} = 3.49$$

*Interpretation:*Reject H_0

Hence there is a significant difference between the means
on the two trials.

In the above example non-directional hypothesis was put forward. The H_1 did not mention any direction of the difference. The difference could be in favour of the first trial also. However, if our hypothesis had been that practice increases test scores and there would be gain on successive trials, a one-tailed test would have been used. In that case the critical value of t from the table would have been read as follows:

- (i) For 9 df, the 0.1 level is read from the .02 col. ($P/2=.01$); t value = 2.82.
- (ii) For 9 df, the .05 level is read from the .10 col. ($P/2=.05$); t value = 1.83.

The calculated value of t in this example is much larger than the value of t required for significance at both the levels with the directional hypothesis. Hence the mean gain from the first to the third trial on the test is significantly larger.

9.7 The Significance of Difference Between Standard Deviations

Independent Samples

When samples are independent, i.e. different groups have been studied or there is evidence that the two tests given to the same group are uncorrelated, the significance of the difference between two σ 's may be found by the formula:

$$SE_{\sigma_1 - \sigma_2, \sigma_{\sigma_1 - \sigma_2}} = \sqrt{\sigma_{\sigma_1}^2 + \sigma_{\sigma_2}^2} \quad (9.5)$$

(SE of the difference in two independent samples)

At times, the psychologists and the educational researchers are interested more in the variability than in the means of the groups. Experimental treatments may be tried out which affect variability. Hence, the need for testing the difference in the standard deviations.

Example: Suppose, on a test of verbal ability, two groups of 32 boys and 50 girls obtain standard deviations equal to 8 and 6 respectively. Is the difference in the two standard deviations significant?

Solution

First calculate the values of σ_1 and σ_2 by Formula (8.9).

$$\sigma_{\sigma_1} = \frac{\sigma_1}{\sqrt{2N_1}} = \frac{8}{\sqrt{2 \times 32}} = 1.00 \text{ (SE for boys group)}$$

$$\sigma_{\sigma_2} = \frac{\sigma_2}{\sqrt{2N_2}} = \frac{6}{\sqrt{2 \times 50}} = .6 \text{ (SE for girls group)}$$

σ_1 and σ_2 are SD's for boys and girls respectively, and so are N_1 and N_2 .

$$\begin{aligned}\sigma_{\sigma_1 - \sigma_2} &= \sqrt{(1.00)^2 + (.6)^2} \\ &= 1.17\end{aligned}$$

$$t = \frac{\sigma_1 - \sigma_2}{\sigma_{\sigma_1 - \sigma_2}} = \frac{8 - 6}{1.17} = \frac{2}{1.17} = 1.71$$

Interpretation

The null hypothesis, $H_0 = \sigma_1 = \sigma_2$, could not be rejected because the observed value of t , 1.71, is less than the critical value of t with $df = 32 + 50 - 2 = 80$, at .05 level. Hence, the boys and girls do not differ significantly on variability on the test of verbal ability.

Correlated Samples

Situations arise when the same group of subjects is to be tested before and after a particular experimental treatment. The two groups, to be compared may be matched or correlation may be introduced due to some other factors. In these situations, formulas for the comparison of SD's for independent samples do not apply. However, the formula given below can be used for the purpose:

$$\sigma_{\sigma_1 - \sigma_2} = \sqrt{\sigma_{\sigma_1}^2 + \sigma_{\sigma_2}^2 - 2r_{12}^2 \sigma_{\sigma_1} \sigma_{\sigma_2}} \quad (9.6)$$

(SE of the difference between two correlated SD's with large N).

in which σ_1 and σ_2 are the SE's of the individual SD's calculated by Formula (8.9) and r_{12} is the value of correlation between

the two sets of scores. The other procedure of calculation and interpretation remains the same as for independent samples above.

9.8 The Significance of the Difference Between two Independent Proportions

Experimental results sometimes require a test of significance of the difference between two independent proportions taken from two randomly drawn samples. Of the N_1 members of the first group, f_1 have the attribute A. Of the N_2 subjects of the second group, f_2 have this attribute. The question arises, do the two proportions, $f_1/N_1 = p_1$ and $f_2/N_2 = p_2$, differ significantly. Can the two samples be regarded random samples from the same population? This requires the computation of SE of the difference between the proportions. The SE of a single proportion is given by Formula (8.10) reproduced below:

$$\sigma_p = \sqrt{pq/N} \text{ in which } p = \text{sample value of a proportion}$$

$$q = 1 - p$$

$$N = N_o. \text{ of cases in the sample}$$

When a comparison of two sample proportions is involved, the SE of the difference is given by

$$\sigma_{p_1 - p_2} = \sqrt{pq \left(\frac{1}{N_1} + \frac{1}{N_2} \right)} \quad (9.7)$$

(SE of the difference between two proportions)

in which p is an estimate based on the two samples combined and $q = 1 - p$. The value of p is obtained by adding up the frequencies of occurrence in the two samples and dividing it by the sum of the two N 's.

$$p = \frac{f_1 + f_2}{N_1 + N_2} \quad (9.8)$$

A t value is then obtained by dividing the difference of the two proportions by $SE_{p_1 - p_2}$.

Example: In school A, 350 students out of 500, and in School B, 250 students out of 300, passed in a public examination. If the two schools have been selected randomly from a

large number of schools in a district, do the two schools differ significantly in terms of their performance in the public examination?

Solution

$$p_1 = \frac{350}{500} = .7$$

$$p_2 = \frac{250}{300} = .83$$

$$\text{Combined proportion, } p = \frac{350 + 250}{500 + 300} = \frac{600}{800} = .75$$

$$q = 1 - p = 1 - .75 = .25$$

SE of the difference by formula (9.7)

$$\begin{aligned} \sigma_{p_1, p_2} &= \sqrt{.75 \times .25 \left[\frac{1}{500} + \frac{1}{300} \right]} \\ &= \sqrt{.19 \times .005} \\ &= \sqrt{.00095} = .031 \end{aligned}$$

$$t = \frac{.70 - .83}{.031} = \frac{.13}{.031} = 4.19$$

The value of t from Table B with $df = (500 + 300) - 2 = 798$, at .05 level = 1.96; and at .01 level = 2.58.

Interpretation: The calculated value of t is larger than the critical value of t at .01 level. Hence it is significant, and thus the null hypothesis ($H_0: p_1 = p_2$) cannot be retained. The two schools differ significantly in their performance in the public examination.

9.9 The Significance of the Difference Between two Correlated Proportions

When the two proportions have been obtained on the same sample of individuals or on matched samples, the paired observations may exhibit a correlation between the two proportions and must be taken care of while using a test of significance of difference between them.

Example: If 400 senior citizens of a big city answer the two questions as below, is the difference between the two proportions of persons saying 'Yes' to both the questions significant?

	YES	NO
1. Do you have any financial problems?	180	220
2. Do you have any emotional problems?	200	200

The data is arranged below in the shape of a 2×2 contingency table. This arrangement is, of course, based on detailed information about paired observations for each individual. One individual may say 'Yes' to both the questions; the second, 'NO' to both the questions; the third 'Yes' to question one, and 'NO' to question two and so on. This data can be tabulated as below in 2×2 contingency tables, showing frequencies and proportions based on frequencies. The cells in tables of frequencies have been marked by capital letters and those for proportions, by small letters whose arrangement may be noted carefully.

		Frequencies		
		Question 1		
		No	Yes	
Question 2	Yes	160 (A)	140 (B)	220
	No	160 (C)	40 (D)	200
		220	180	400

Proportions

Question 1

	No	Yes		
Q. 2	YES	.15 (a)	.35 (b)	.50
	No	.40 (c)	.10 (d)	.50
		.55	.45	1.00

The null hypothesis is that the proportions of 'Yes' responses to question 1 and question 2 do not differ, except beyond chance. The standard error of the difference between two correlated proportions is given by the formula

$$\begin{aligned}
 SE_{p_1-p_2} &= \sqrt{\frac{a+d}{N}} \\
 &= \sqrt{\frac{.15+.1}{400}} = .0125
 \end{aligned}
 \tag{9.9}$$

The value of $z = \frac{p_1 - p_2}{SE_{p_1-p_2}}$

According to the problem, p_1 and p_2 are

$$180/400 = .45 \text{ and } 200/400 = .50$$

$$\text{Hence } z = \frac{.45 - .50}{.0125} = \frac{.50}{.0125} = 4.00$$

The critical value of z at .05 and .01 levels are 1.96 and 2.58 respectively (Consult Table A in appendix). Hence, the calculated value of z is significant at .01 level.

In case frequencies are to be used, the following formula will apply:

$$z = \frac{D - A}{\sqrt{A + D}} \quad (9.10)$$

(Formula for comparison of Frequencies)

Another general formula for the calculation of $SE_{p_1-p_2}$ is

$$\sigma_{p_1-p_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2 - 2r_{p_1 p_2} \sigma_{p_1} \sigma_{p_2}} \quad (9.11)$$

(SE of difference between two correlated proportions, based on calculation of r)

For this purpose, r between the two percents is given by the phi coefficient, a ratio equivalent to the correlation coefficient in 2×2 tables. Calculation of phi coefficient will be shown in chapter on correlational techniques. σ_{p_1} and σ_{p_2} are SE's of p_1 and p_2 computed according to Formula (8.10).

9.10 The Significance of the Difference Between two r 's

Situations arise in which two correlation coefficients are to be compared. There could be correlations between attitude towards a particular course and achievement in that course for two different groups, say, male and female students. The null hypothesis, $H_0: \rho_1 = \rho_2$ or $H_0: \rho_1 - \rho_2 = 0$. The procedure of calculating an SE of the r has already been explained in a section in the preceding chapter. The comparison of the r 's involves the following steps:

1. Convert the two r 's into Fisher's z function.
2. The $SE_{z_1-z_2} = \sqrt{\sigma_{z_1}^2 + \sigma_{z_2}^2} = \sqrt{\left(\frac{1}{N_1-3} + \frac{1}{N_2-3}\right)} \quad (9.12)$

(Comparison of two r 's through z conversion)

N_1 and N_2 are No. of cases in two groups.

$$3. \text{ Value of } t = \frac{z_1 - z_2}{SE_{z_1 - z_2}} \quad (9.13)$$

The above steps will be demonstrated through an example.

Example: The two correlation coefficients in the illustration given above are .64 and .78 with $N_1=103$ and $N_2=63$. Is there a significant difference in the two r 's.

Solution

From Table C, Fisher's z equivalent of r 's of .64 and .78 are .76 and 1.05 respectively. By formula (9.12):

$$SE_{z_1 - z_2} = \sqrt{\frac{1}{103-3} + \frac{1}{63-3}} = .164$$

$$t = \frac{1.05 - .76}{.164} = \frac{.29}{.164} = 1.77$$

The $df = N_1 + N_2 - 3 = 103 + 63 - 3 = 163$, and critical value of t at .05 level is 1.97. The observed value of t is smaller than the critical value of t . Hence, the H_0 cannot be rejected. There is no significant difference between the two correlations.

In case where the same sample has been tested on more than one variables and the correlations between the variables have been computed, the situation involves an SE of the correlated samples. Suppose a group of 10 randomly selected students have been tested on their attitude towards Mathematics, their achievement in mathematics and general mental ability. The three pairs of correlations are: r_{12} , r_{13} and r_{23} with their values, .8, .7 and .2 respectively. If we wish to compare r_{12} and r_{13} ; or r_{12} and r_{23} ; or r_{13} and r_{23} the procedure mentioned above does not apply because it is based on correlated samples.

Hence, to test the difference between r_{12} and r_{13} , we use

$$t = \frac{(r_{12} - r_{13}) \sqrt{(N-3)(1+r_{23})}}{\sqrt{2(1-r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}} \quad (9.14)$$

(Difference in two r 's in correlated samples)

Inserting the values, we obtained

$$t = \frac{(.8 - .7) \sqrt{(103 - 3)(1 + .2)}}{\sqrt{2(1 - .8^2 - .7^2 - .2^2 + 2 \times .8 \times .7 \times .2)}}$$

$$= \frac{.1 \times 10.06}{.33} = 3.21$$

Interpretation

With $df = 103 - 3 = 100$, a t of 2.63 is required for significance. Since the observed value is higher than 2.63, it is significant at .01 level. Hence, there are significant differences in the two r 's and the H_0 cannot be retained.

9.11 Two-Tailed and One-Tailed Tests of Significance

The null hypothesis denotes that the difference between the obtained means may be either plus or minus and as often in one direction as in the other from the true (population) difference of zero. Hence for determining probabilities, both the tails of the sampling distribution are used. (See Figure 9.1). When the primary concern is with the direction of the difference rather than with its existence in absolute terms, the situation calls for a one-tailed test and only one-tail of sampling distribution is used. (See example 3). The procedure for picking up the values of the critical ratio, t , for a particular significance level under one-tailed, test are as follows:

<i>Level of significance</i>	<i>Column of the t table to be consulted</i>	<i>Reason</i>
.05	.10	$P/2 = .05$
.01	.02	$P/2 = .01$

The t tables given in books on statistics are generally meant for a two-tailed situation and the probability is distributed equally in the two tails (For .05, $P = .025$ in each of the two tails; and For .01, $P = .005$ in each of the two tails). Hence, the probability must be doubled to obtain the correct value of the critical ratio in a two-tailed test.

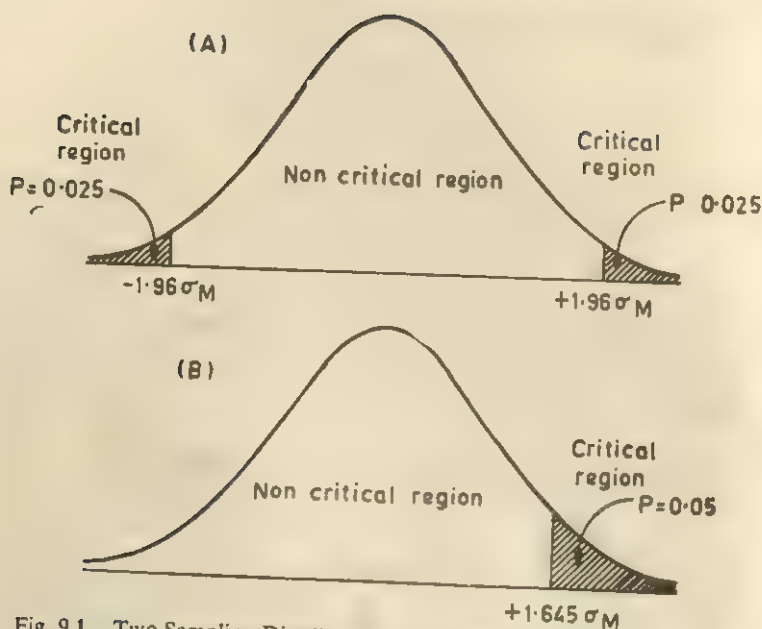


Fig. 9.1. Two Sampling Distributions of Means showing the Critical Regions and Non-Critical Regions in (A) Two-tailed or Non-directional Test and (B) One-tailed or Directional Test.

9.12 Type I and Type II Errors

Research requires testing of hypothesis. In this process, two types of wrong inferences can be drawn. These are called Type I and Type II errors.

Type I error is committed when we reject a null hypothesis by marking a difference significant, although no true difference exists.

Type II error is committed when we accept a null hypothesis by marking a difference not significant when a true difference actually exists. These can be diagrammatically shown as below:

	Decision	
	Reject H_0	Accept H_0
H_0 True:	Type I Error	No Error
H_0 false:	No Error	Type II Error

Exercises for Practice

- 9.1 Two groups of students selected randomly from two different colleges, were administered an attitude scale, and the following data was collected:

	N	M	SD	Do the two groups differ significantly in their attitudes?
College: I	40	30.5	6.0	
College: II	50	25.4	5.0	

- 9.2 A group of 10 students was given four trials on a test of physical efficiency. Their scores on the I and IV trials are given below. Test whether there was a significant gain from the first to the fourth trials.

Students	Trial-I	Trial-IV
1	15	20
2	16	22
3	17	22
4	20	25
5	25	35
6	30	30
7	17	21
8	18	23
9	10	17
10	12	20

- 9.3 A group of 40 students was administered a test of intelligence twice. The data so collected is given below: Test whether there was a significant difference in the means on the two testings.

	M	SD
Testing-I	25	8
		$r = .65$
Testing-II	35	5

- 9.4 Define the following as precisely as possible:

- (i) Level of significance
- (ii) Correlated samples
- (iii) Standard Error of the Mean
- (iv) Sampling Distribution

- 9.5 (a) Given two random samples of size 450 and 550 respectively each with sample values $p_1=.66$ and $p_2=.50$. Test the significance of the difference between p_1 and p_2 .
(b) Test the difference between $p_1=500/600$ and $p_2=150/200$ (Independent samples).
- 9.6 Given two independent samples of size 200 each with $\sigma_1^2=121$ and $\sigma_2^2=400$, test the hypothesis that the variances are significantly different from each other.
- 9.7 If the groups are correlated, and the $\sigma_1=15$ and $\sigma_2=10$ and $r=.6$, with $N=100$, test the hypothesis that the standard deviations are significantly different.
- 9.8 (a) Calculate values of z_r for $r=.60$; $r=.55$; $r=.25$; and $r=-.85$.
(b) Why is the conversion of r 's into z_r 's required in calculating SE of r ?
(c) Who devised this method of conversion?
- 9.9 Compare the following pairs of r 's for significance of differences. Use .05 level of significance.
(a) $r_1=.60$; $r_2=.40$; $N=84$
(b) $r_1=.80$; $r_2=.00$; $N=83$
(c) $r_1=.20$; $r_2=.70$; $N=124$
(d) $r_1=.25$; $r_2=.55$; $N=403$
(e) $r_1=.62$; $r_2=.00$; $N=102$
- 9.10 Three psychological tests are administered to a sample of 100 students. The correlations obtained are $r_{12}=.80$; $r_{13}=.50$; and $r_{23}=.40$. Is r_{12} significantly different from r_{13} ?

CHAPTER 10

THE CHI-SQUARE TEST AND OTHER NON-PARAMETRIC METHODS

Problems in social research frequently involve the counting of a number of persons, objects or responses as they occur under various categories of classifications. For example, school children may be classified and counted according to their reading ability, mathematical ability, or their modes of behaviour. Adult citizens may be classified according to whether they are "in favour of", "indifferent to", or "opposed to" a particular social reform. The Chi-square test is suitable for analyzing data and problems like those mentioned above. When data are in the form of discrete categories and frequencies, Chi square is, perhaps, the most suitable test to compare the obtained set of *observed* frequencies in given categories with a set of *theoretical* or *expected* frequencies occurring within them. The number of categories may be two or more and the theoretical frequencies may be determined in a number of different ways depending upon the nature of the problem under consideration. *Chi-square is the statistic which measures the "divergence" of fact from hypothesis in the sample at hand.* The Chi square, needed to test this and a wide variety of similar hypotheses, may be defined as

$$\chi^2 = \frac{\sum (f_o - f_e)^2}{f_e} \quad (10.1)$$

in which, f_o = Observed frequency in a single category
 f_e = Expected, theoretical or hypothetical frequency
 Σ = Sum of

From the formula, it is evident that Chi-square is an index of the divergence of fact from hypothesis. If each of the

observed frequency agreed exactly with the corresponding theoretical frequency, the value of χ^2 would be zero. The larger the divergence between the observed and theoretical frequency, the larger the value of Chi-square. Chi-square is based on the squares of the deviations, $(f_o - f_e)^2$, and hence does not take the direction of the deviations into account. This is a limitation of the Chi-square.

The form of sampling distribution of Chi-square depends only upon the degrees of freedom in the table from which Chi-square has been calculated. In other words, Chi square shows the same distribution for all random samples in which the number of degrees of freedom is the same, regardless of the size of the sample (so long it is fairly large, say 50 or more, and no theoretical frequency is very small, say 10 or less). Karl Pearson gave the concept of Chi-square and worked out its sampling distribution on the basis of the following:

$$y = y_0 e^{-\frac{1}{2} \chi^2} (\chi^2)^{(df-1)/2} \quad (10.2)$$

Chi-square tables (Table F) have been prepared on the basis of this equation. Figure 10.1 shows the distribution of Chi-square for different degrees of freedom.

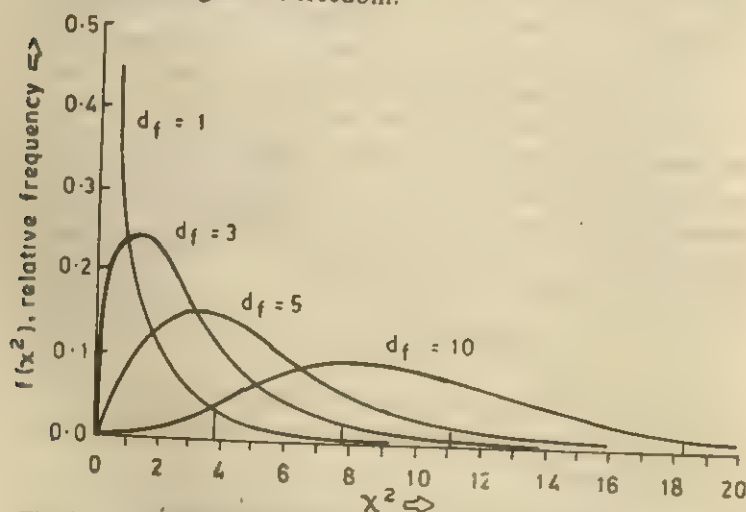


Fig. 10.1. Chi-Square Distribution and 5 per cent Critical Regions for various Degrees of Freedom.

The value of Chi square is always positive, a circumstance which results from squaring the difference between observed and theoretical frequencies. Values of χ^2 range from 0 to infinity. The right hand tail of the curve is asymptotic to the ordinate as well as to the abscissa. This statistic is used in tests of significance in much the same way as the normal distribution, t distribution or F distribution.

10.1 Degrees of Freedom, df

The number of degrees of freedom in a table of frequencies is the number of those frequencies to which we may assign arbitrary values and still satisfy the external requirements in terms of row and column totals. If we consider each frequency as occupying a cell in the contingency table, the degrees of freedom is the number of cells that may be filled at will. For example, we may have a 2x2 contingency table, like the following:

(a)	(b)	
10	20	30
(c)	(d)	
15	25	40
25	45	

The restriction imposed is that the cell frequencies in each row and column must add up to a fixed total for that row or column. We may note here, that the frequencies of only one cell can be varied at will. Others get fixed immediately to make up the row and column totals. Suppose we change f of 10 to 12. The frequencies of cell (b) and cell (c) will be fixed to 18 and 13 respectively to conform to the marginal totals of 30 and 25. The frequencies in cell (d) will also get fixed to make up the marginal totals. Hence in this table, we have only one

degree of freedom. In 2×2 or larger contingency tables, the df may be calculated as below:

$$df = (r-1)(c-1) \quad (10.3)$$

(Degrees of freedom in contingency tables)

where, r =number of rows; and c =number of columns.

In a 3×3 contingency table, $df = (3-1)(3-1) = 4$.

There are other types of restrictions that might be imposed on a table besides those concerned only with totals. An illustration of these will be given in a test of goodness of fit for normality. The steps involved in the use of the Chi square are as follows:

1. State the null hypothesis, H_0 , and the alternative hypothesis, H_1 .
2. State the level of significance (α) and the sample size (n).
3. Determine the critical region based on df and state the decision rule. Critical value of Chi square will be found from the Chi square table (Table F).
4. Compute the value of Chi-square by using formula (10.1).
5. Take a decision to reject, or not to reject the null hypothesis.

We shall take a few illustrations to clarify the process.

10.2. Test of the Hypothesis of Equal Probability

Example: Sixty Post-graduate students were asked to express an opinion on the issue "Should India make an atomic bomb?" by marking on a three point scale—"Yes", "?", "No". Thirty of the group marked "Yes"; 12 "?"; and 18 "No". Do these results indicate a trend significantly different from equal probability of opinion in each of the three categories?

The solution is given in Table 10.1

TABLE 10.1

Computation of Chi-square test of hypothesis of equality (Example about atom bomb)

Hypotheses

$$H_0 = f_{yes} = f_{?} = f_{No}$$

$$H_1 = f_{yes} \neq f_{?} \neq f_{No}$$

Decision Rule

Given: $\alpha = .05$; and a 3×2 contingency table with

$$df = (r-1)(c-1) = (2-1)(3-1) = 2$$

If $\chi^2_{obs} < 5.99$, accept H_0

If $\chi^2_{obs} \geq 5.99$, reject H_0

Computation

The data is arranged in a contingency table

	Responses			
	Yes	?	No	
Observed (f_o)	30	12	18	60
Expected (f_e)	20	20	20	60
$(f_o - f_e)$	10	8	2	
$(f_o - f_e)^2$	100	64	4	
$\frac{(f_o - f_e)^2}{f_e}$	5	3.2	.2	

$$\chi^2 = \frac{\sum (f_o - f_e)^2}{f_e} = 8.4$$

Interpretation

Reject H_0 . The responses indicate a trend significantly different from equal probability.

10.3. Test of Hypothesis of Independence (Difference)

Chi square can be used to test whether two variables or attributes were *independent* or unrelated. Suppose we wish to know whether the variable of internal-external control was independent of sex of the subjects. We have three categories of subjects on internal-external control and two categories of subjects on sex. The obtained frequencies are shown below. Independence values or expected frequencies have also been given in parentheses. Each cell has been marked by a letter for identification purposes only.

		Externality			
		High (H)	Middle (M)	Low (L)	
Boys	(a)	20 (12)	(b) 10 (12)	(c) 10 (16)	40
Girls	(d)	10 (18)	(e) 20 (18)	(f) 30 (24)	60
		30	30	40	

Steps of procedure are given below:

Steps

1. *State the Hypotheses:* The null hypothesis H_0 is:

The proportion of boys in three alternative categories of externality is the same as the proportion of girls in these three categories or externality is independent of sex. The statistical alternative hypothesis (H_1) is: The proportion of boys in the three alternative classifications of externality is not the same as that of the girls in these categories or the externality is related to sex.

In symbolic form

$$H_0 =: P_{BH} = P_{BM} = P_{BL}$$

$$P_{GH} = P_{GM} = P_{GL}$$

$$H_1 : P_{BH} \neq P_{BM} \neq P_{BL}$$

$$P_{GH} \neq P_{GM} \neq P_{GL}$$

2. *Decision Rule:* Given: $\alpha = .05$ and $df = (3-1)(2-1) = 1$

If $\chi^2_{obs.} < 5.99$, accept H_0

If $\chi^2_{obs.} \geq 5.99$, reject H_0

3. *Computation:* Computation of independence values or expected frequencies for each cell may be found by multiplying the marginal totals common to a particular cell and then dividing this product by the grand total of frequencies in the table. The calculation of the same is shown below; The results are then shown in parentheses in each cell.

Calculation of expected frequencies or independence values for various cells:

$$(a) \frac{40 \times 30}{100} = 12; (b) \frac{40 \times 30}{100} = 12; (c) \frac{40 \times 45}{100} = 18,$$

$$(d) \frac{60 \times 30}{100} = 18; (e) \frac{60 \times 30}{100} = 18; (f) \frac{60 \times 40}{100} = 24.$$

$$4. \text{ Chi square: } \frac{(20-12)^2}{12} + \frac{(10-12)^2}{12} + \frac{(10-16)^2}{16} + \frac{(10-18)^2}{18} + \frac{(20-18)^2}{18} + \frac{(30-24)^2}{24} \\ = 5.33 + .33 + 2.25 + 3.56 + .22 + 1.50 = 13.19$$

5. *Interpretation:* Reject H_0 and accept H_1 . It shows that the distribution of externality is not independent of sex. It may further be concluded that the distribution of externality differs significantly between the two sexes.

10.4 Test of the Hypothesis of Normality

Chi square can be used to test significance of divergence of the observed results from those expected on the hypothesis of a

normal distribution. The hypothesis may be set up in such a way that it asserts that the observed frequencies of an event follow a normal distribution instead of being equally probable.

Example: Fifty students were rated on "aggressiveness" by a group of their teachers. The ratings were done by consensus on a three fold classification: generally aggressive (10 students); sometime aggressive (28 students), and seldom aggressive (12 students). If aggressiveness is presumed to be normally distributed in this population of students, does this distribution of ratings differ significantly from normality?

Solution of the numerical problem is given hereunder:

Steps of Procedure

1. Hypotheses

H_0 : The observed frequencies are normally distributed.

H_1 : The observed frequencies are not normally distributed.

2. Decision Rule

Given $\alpha = .05$; and a 2×3 contingency table with $df = (2-1)(3-1) = 2$.

If $\chi^2_{obs} < 5.99$, accept H_0 .

If $\chi^2_{obs} \geq 5.99$, reject H_0 .

3. Computation

	Generally Aggressive	Sometime Aggressive	Seldom Aggressive	
Observed (f_o)	10	28	12	50
expected (f_e)	8	34	8	50
$(f_o - f_e)$	2	6	4	
$(f_o - f_e)^2$	4	36	16	
$\frac{(f_o - f_e)^2}{f_e}$.50	.94	2.00 = 3.44	
				$= \chi^2$

In the contingency table above, entries in row 1 give the number of students classified in each of the three categories. Entries in row 2 have been calculated with the help of a table of area under the normal curve (Table A) given in the appendix. The baseline of the normal curve (taken to extend over 6σ) has been divided into segments of 2σ each and the proportions enclosed by these limits found.

Segment	Proportion	Frequencies out of 50
Between $+3\sigma$ and $+1\sigma$.16	$.16 \times 50 = 8$
Between $+1\sigma$ and -1σ	.68	$.68 \times 50 = 34$
Between -1σ and -3σ	.16	$.16 \times 50 = 8$

These frequencies have been entered in the contingency table in the second row against *fe*.

Other steps of computation are self evident.

4. *Interpretation:* The value of observed Chi square is less than the critical value of Chi-square required for significance at .05 level and with $df = 2$. Hence, accept H_0 . It may be concluded that the distribution of observed frequencies is not significantly different from normality.

Problems on normal distribution requiring larger classifications can also be tackled in the same manner. (Also consult Chapter on Normal distribution).

10.5 Calculation of Chi Square for 2×2 Tables

A frequently occurring type of contingency table is the 2×2 or four-fold contingency table. The value of the Chi square for the test of independence can be readily obtained for such a table without calculating the expected values. The cell and marginal frequencies in the Table be designated as below:

A	B	A + B
C	D	C + D
A + C	B + D	

Chi square can be calculated by using the formula

$$\chi^2 = \frac{N(AD - BC)^2}{(A+B)(C+D)(A+C)(B+D)} \quad (10.4)$$

in which,

N = Total Number in the table

A, B, C, D = Frequencies in each of the four cells

AD, BC = Cross products of the cell frequencies $A \times D$ and $B \times C$ respectively.

$A+B, C+D, A+C, B+D$ = Marginal totals of frequencies.

Thus, the formula may read, "Chi square is equal to N times the square of the difference of the cross products divided by the product of the four marginal totals."

Example: In the following table, 55 students have been placed in four cells according to their teachers' ratings as good performers and poor performers, and success/failure on an intelligence test item. Is the performance independent of the intelligence?

		Test item		
		Fail	Pass	
Good Performers	(A)	10	15	25 (A + B)
	(C)	16	14	30 (C + D)
Poor Performers		26 (A + C)	29 (B + D)	55 (N)

$$\chi^2 = \frac{55[(10 \times 14) - (15 \times 16)]^2}{25 \times 30 \times 26 \times 29} = .97$$

The critical value of Chi square with $df=1$, on $\alpha=.05$, is 3.84. The observed value of Chi square is not significant. Hence, the data do not provide enough evidence that the test item differentiates between individuals on the basis of their performance (as rated by their teachers).

When entries in a four-fold table are quite small (for example, 5 or less), Yates' correction for continuity should be applied. Formula (10.4) then becomes:

$$\chi_c^2 = \frac{N(|AD-BC| - N/2)^2}{(A+B)(C+D)(A+C)(B+D)} \quad (10.5)$$

in which, $|AD-BC|$ means the absolute difference of the two cross products. Other symbols, are as explained above. Applying the formula to the data given above, we have

$$\chi_c^2 = \frac{55(|140-240| - 55/2)^2}{25 \times 30 \times 26 \times 29} = .51$$

This value is smaller than that of the uncorrected Chi square. Yates' correction will always reduce the size of the Chi square. Its effect is more crucial when entries are small. When χ^2 is marginally significant, the χ_c^2 may well fall below the level set for significance. However, if χ^2 is already not significant, χ_c^2 will be even less so.

10.6 Yates' Correction for Continuity

Chi square is subject to considerable error in the following two situations and requires the use of Yates' Correction for Continuity:

1. *When working with 2×2 tables.* Chi square is based on the assumption that adjacent frequencies are connected by a continuous and smooth curve (like the normal curve) and are not discrete numbers. In 2×2 tables this continuity of the curve is broken and needs a correction.

2. *When entries are very small.* Chi square is not stable like any other statistic based on probability, when computed from a table in which any experimental frequency (f_o) is less than 5. Hence a correction of continuity is required.

Failure to use the correction causes the probability of a given result to be greatly underestimated and the chances of its being called significant considerably increased.

Yates' correction for continuity requires subtracting of .5 from each $(f_o - f_e)$ difference as shown in the following example :

	Right	Wrong
f_o	8	2
f_e	5	5
$(f_o - f_e)$	3	3
correction $(-.5)$	2.5	2.5
$(f_o - f_e)^2$	6.25	6.25
$\frac{(f_o - f_e)^2}{f_e}$	1.25	1.25 = 2.50 = χ_c^2

falls far below the .05 significance level which requires a critical value of 3.841. However, the uncorrected value of χ^2 , if calculated comes to 3.6. which approaches nearer to the critical value.

10.7 Chi Square from Percentages

Chi square test can be used when table entries are in terms of percentages or proportions. However, in such cases, a correction for size of the sample becomes essential. It is because of the fact that percentage does not indicate whether the actual frequency from which percentage has been calculated was large enough. A frequency of 6 out of 10 gives the same percentage as a frequency of 60 out of 100. However, the latter is more significant. The calculation of Chi square from percentages is shown with the help of an example below. N in this case is equal to 10.

	Right	Wrong	
fo	80%	20%	100%
fe	50%	50%	100%
(fo-fe)	30%	30%	
correction (-5%)	25%	25%	
(fo-fe) ²	625	625	
$\frac{(fo-fe)^2}{fe}$	12.5	12.5	
$\chi^2\%$	25		
χ^2	$= \chi^2\% \times N/100 = 25 \times 10/100 = 2.5$		

The $\chi^2\%$ based on percentages must be brought back to its proper value in terms of original numbers by multiplying it by $N/100$ as shown above.

10.8 General Observations on Chi Square

10.8.1 Assumptions of the Chi square test

Chi square test is based on the following assumptions:

- (i) The two samples are independent from one another. This implies that different and unrelated sets of subjects are selected.
- (ii) The subjects within each group must be randomly and independently sampled.
- (iii) Each observation must qualify for one and only one category. It implies that f's in each cell must be distributed in mutually exclusive categories.
- (iv) The sample size must be relatively large.

The two important assumptions of normality of distribution and homogeneity of variance' which underly all important parametric tests do not apply to Chi square.

10.8.2 One-tailed and two-tailed situations

Tables of Chi square used for significance are based on one tailed-tests only, using the tail to the right of the sampling distribution of χ^2 . Although one tail only of the sampling distribution of χ^2 is used, the table values are those required for testing the significance of a difference regardless of direction, i.e., for two-tailed tests. This is because of the fact that, in effect, χ^2 is the square of the normal deviate for 1 degree of freedom, thus incorporating both tails of the normal curve in the right tail of the χ^2 curve. In many situations where Chi square is applied, the idea of a directional, or one tailed test has little meaning. In tests of independence and goodness of fit, we are usually not concerned with the direction of the difference observed. However, if a one tailed test is required, the proportionate areas in the Chi square table should be halved.

10.8.3 Reduction of an $R \times C$ table to a 2×2 table

Sometimes a table with more than two rows and more than two columns may be reduced to a 2×2 table by combining the tail frequencies. The procedure is statistically legitimate provided the points of division (dichotomy) of the two variables are picked up independently of the cell frequencies and not for maximizing the association in the data to obtain a significant Chi-square.

10.8.4 Additivity of Chi-square

In many experimental studies, Chi square from different samples may be added to provide an overall test of hypothesis. In such cases, the df for the new Chi-square will be the sum of the separate df's. Each repetition of the experiment should be done on samples of almost equal size and drawn independently and at random. Suppose a sex difference with respect to an attitude to a certain question is expected. A test is run separately in grades 10, 11 and 12 for the purpose and the separate Chi-squares are found to be not significant. The Chi squares for the three classes may be combined to test an overall hypothesis covering all the three classes with greater likelihood of obtaining a significant result.

The applications and uses of Chi square table described in this chapter are fundamental and elementary ones generally applicable to simple research problems. Several tests which are required in more complex situations make use of Chi square to test significance. Some such tests are comparable to those of analysis of variance, repeated measures, trend analysis, matched pairs, etc. These tests make up a whole separate area of study known as non-parametric or distribution-free statistics.

10.9 Non-Parametric Statistical Tests

Most of the tests of significance so far described in previous chapters are referred to as parametric tests because these involve the estimation of at least one population value or parameter. For example, in F test a population variance estimate is needed as the error term. The tests to be described in this chapter are sometimes called non-parametric because they do not involve the estimation of any parameter on population value. Moreover, the assumption of normality of distribution which is a pre-requisite in such tests as F, t and Z, is also not made. Therefore, the techniques presented in this chapter are sometimes called *distribution-free* meaning thereby that in the application of these techniques the population distribution need not necessarily be normal.

Non-parametric tests are generally preferred to parametric tests in the following situations:

1. Whenever there are doubts that the distribution is not normal or the non-normality has been established through some statistical procedures. When samples are not from a normal distribution, the use of a normal theory test with level of significance, α , does not assure that the probability of an error of the first kind is controlled at level α . Such probability is indeterminate because generally there is no way of knowing the direction and degree of departure from α .

On the other hand, if a non-parametric test is used with level of significance α then for any parent population the probability of an error of the first kind is actually equal to α or less than α because of discreteness.

2. When measurements are in terms of nominal and ordinal scales, the use of non-parametric tests will be more appropriate.

A nominal scale does not involve any measurement as such and classifies individuals into categories that are qualitatively different with no ordering implied. The categories may be assigned numerical designations and number of persons in each category expressed as proportion or percentage. Ordinal measurements are in terms of rank order based on qualitative or quantitative considerations. The level of measurement and type of data available do restrict the type of statistical techniques that can be used. Chi-square test can be used for handling nominal data and data in terms of ranks can be handled through Spearman's 'rho or Kendall's tau or the tests to be described in this book

10.9.1 Sign Test

The sign test is the simplest of all distribution-free statistics and carries a very high level of general applicability. It is applicable in situations in which the critical ratio, t , test for correlated samples cannot be used because the assumptions of normality and homoscedasticity are not fulfilled. The students are aware of the fact that certain conditions in the setting of the experiment introduce the element of relationship between the two sets of data. These conditions generally are a pre-test, post-test situation; a test re-test situation; testing of one group of subjects on two tests; formation of 'matched groups' by pairing on some extraneous variables which are not the subject of investigation, but which may affect the observations.

Suppose an experimenter selects a group of 20 persons and divides them into two groups, control and experimental, by matching pairs of subjects on intelligence. One of the groups is then randomly assigned to a leadership training camp. The control group receives no such training. On completion of the camp, independent observers rate the leadership qualities of each of the 20 subjects on a 30-point scale. The rating scale is an extremely crude one and the assumption of equality of intervals and the shape of the distribution of scores cannot be safely made. The only assumption we can safely make is that any difference between two paired scores is a valid indicator of the direction and not the magnitude of the differences.

The data have been arranged in Table 10.2. The null hypothesis to be tested is that the probability that any difference will be positive is equal to the probability that it will be negative. Since it is a two category population of differences (Positive differences and Negative differences) H_0 can be expressed in precisely the same manner as in the binomial test with $P=Q=.5$.

Hypotheses: $H_0 : P_{negative} = P_{positive} = \frac{1}{2}$

$H_1 : P_{negative} \neq P_{positive} \neq \frac{1}{2}$

Decision Rule:

If $P_{obs} > .05$ Accept H_0

If $P_{obs} < .05$ Reject H_0

TABLE 10.2

Ratings of two Groups on Leadership Qualities

Matched pair	Experimental	Control	Sign of difference
1	20	18	+
2	15	10	+
3	25	21	+
4	28	17	+
5	10	12	-
6	18	10	+
7	11	6	+
8	9	8	+
9	7	7	0
10	14	16	-

Number of zero differences = 1

Number of negative differences = 2

Number of positive differences = 7

There are 10 differences in all, 7 are plus, 2 are minus and one is zero. Since the zero differences are neither plus nor minus, they are excluded from N as well as either of the two categories of + and - signs. In our example, we now have 9 pairs out of which 7 are positive (or 2 are negative). To test the hypothesis we shall first expand the binomial $(p + q)^9$, in which the probability of a + is denoted by p and the probability of a - by q .

$$(p + q)^9 = p^9 + 9p^8q + 36p^7q^2 + 84p^6q^3 + 120p^5q^4 + 126p^4q^5 + 84p^3q^6 + 36p^2q^7 + 9pq^8 + q^9 \quad (10.6)$$

The total number of combinations is $2^9 = 2^9 = 512$. Now by adding the numerical coefficients of the first three terms (namely, $p^9 + 9p^8q + 36p^7q^2$) the number of combinations which contain 7 or more plus signs out of 9 is obtained which is 46. The probability, under the binomial, of the occurrence of 7 or more + signs out of 9, can now be determined by dividing the number of combinations having 7 or more + signs by the total number of combinations. In our example, it is $46 \div 512 = .0898$ or .09 (rounded to two decimal places). For convenience this value can also be obtained by referring to Table G in Appendix in which cumulative probabilities under the binomial are given. Entering Table G with $N=9$ and x , the number of signs having smaller value, i.e. - signs, $x=2$, the probability given is .090 which checks with the one calculated above.

It is a one tailed probability. The two tailed probability then, is $2 \times .09$ i.e., .18. Using $\alpha = .05$ for two tailed test, we accept H_0 because $p > .05$. For one tailed test also, the H_0 cannot be rejected because $p > .10$.

* The probability of a particular event can be obtained by using the formula $n! / (r!p^r q^{n-r})$. The numerical coefficient of a term can be calculated by using the first term of the formula. Here the numerical coefficient of

the third term in the numerical example, $9C_7 = \frac{9!}{2!7!} = 36$, since $p=q=1/2$, the formula reduces to $(36)(1/2)^9(1/2)^2$ or $36/512$. In the same manner, probability of occurrence of 8 + and 9 + signs can also be obtained, which calculated on the data above, are $9/512$ and $1/512$ respectively. The cumulative probability of the occurrence of 7 or more plus signs can, then, be calculated by adding the three probabilities $36/512 + 9/512 + 1/512 = 46/512 = .089$.

10.9.2 Sign Test with Large Samples

When the number of pairs is 20 or more an approximation to the exact probabilities can be obtained by calculating χ^2 corrected for continuity by using a 2 × 2 contingency table. The theoretical frequencies in each cell will be $N/2$ (N here does not exclude the pairs having zero differences). The example solved in the previous section though has $N = 20$, yet is repeated here for demonstrating the technique and the computational steps involved in it.

	+	-
$f_{observed} (f_o)$	2	3
$f_{theoretical} (f_e)$	5	5
$f_o - f_e$	2	3
$- .05$	1.5	2.5
$(f_o - f_e)^2$	2.25	6.25
$\frac{(f_o - f_e)^2}{f_e}$.45	1.25, $\chi^2 = 1.70$

For a two-tailed test and $df = 1$, this value is significant at a .18 approx. Which is a good approximation to the value found in the previous section. In cases involving $df = 1$, $z = \chi^2 / 1$ (10.7). Applying this formula to the problem above, we have

$$z = \chi^2 / 1.70 \quad \text{or } z = 1.004$$

Referring to a normal probability table, we find that for a two-tailed test this is significant at a .19 approximately. The difference in this value and the values obtained earlier can be attributed to the smallness of N .

Walker and Lev (1958) report another procedure for calculating χ^2 corrected for continuity by using the following formula and regarding it as a normal deviate

$$z = \frac{2m \pm 1}{\sqrt{N}} - \sqrt{N} \quad (10.8)$$

(m stands for the number of signs smaller in number).

The 1 in the numerator is added or subtracted in such a way as to change $2m$ to a value nearer N . Thus, in our example, $m=2$, $N=9$, we should have

$$z = \frac{2(2)+1}{\sqrt{9}} - \sqrt{9} = 1.33$$

This provides a very close approximation to the value of z computed previously.

Assumptions of Sign Test

The assumptions underlying the sign test are:

- (i) The differences are continuously distributed, and
- (ii) The differences are independent of each other.

As stated earlier, no assumption regarding the form of the distribution or homogeneity of variance is required.

Some other uses of the Sign Test

The Sign Test can be applied to test certain more general hypotheses. It involves an additional assumption that the measurements are on a scale of equal units. By subtracting a constant C , from each difference ($\text{Difference} = X_{Ai} - X_{Bi} - C$) the null hypothesis is that the median difference $X_A - X_B$ in the population is at least C . The hypothesis is rejected if too many differences are positive.

If the data fulfil another assumption i.e., of a zero point on the measurement scale, the difference can be calculated as follows:

$$\text{Difference}_i = X_{Ai} - KX_{Bi} \quad (10.9)$$

and the hypothesis

$$H_0 = P(X_A > KX_B) \leq .5$$

can be tested.

As it is clear from the above two assumptions—equal units and a zero point—these extensions of the sign test can be used only with continuous variables.

10.9.3. The Median Test

It compares the median of two independent samples and is a non-parametric replacement of its parametric t test for comparing the means of two independent samples. It provides a procedure for testing whether two independent samples differ in central tendencies. In other words, the use of the median test allows for testing whether two independent groups have been drawn from populations with the same median. The test can be used for unequal groups also. The hypothesis to be tested can be stated either in non-directional (two tailed test) or directional terms (one tailed test). The pre-requisite of the test is measurements at least on an ordinal scale. The computation requires the following steps of procedure:

1. Determine a common median for both the groups combined.
2. Dichotomize both sets of scores separately at the common median and put the data in a 2×2 contingency table as given below:

	Group I	Group II	Total
No. of scores above combined median	A	B	A + B
No. of scores below combined median	C	D	C + D
	A + C	B + D	N = $n_1 + n_2$

The rationale of the test is that if both Group I and Group II are samples from populations whose median is the same, we would expect about half of each group's scores to be above the combined median and about half to be below the combined median.

3. Use the Chi square test*

*Use Fisher Test, instead of Chi square test, if (i) $n_1 + n_2 < 20$, or (ii) when any cell has $f_e < 5$ although $n_1 + n_2$ is between 20 and 40.

4. If the P yielded by the test is equal to or smaller than α , reject H_0 .
5. If some scores fall at the combined median then (i) drop them from analysis if they are only a few and $n_1 + n_2$ is large, or (ii) dichotomize the groups as those scores which exceed the median and those which do not. The troublesome scores may be included in the second category. The steps are illustrated in Table 10.3.

TABLE 10.3

Computation of a Median Test for Two Samples

Hypotheses: H_0 : No. of scores below Combined Median in Sample I = No. of scores below combined median in Sample II

H_1 : No. of scores below combined median in sample I \neq No. of scores below combined median in Sample II

Decision Rule: given $\alpha = .05$, $n_1 + n_2 > 20$ and no $f_e < 5$; $df = 1$

If $\chi^2_{obs.} < 3.84$, Accept H_0

If $\chi^2_{obs.} \geq 3.84$, Reject H_0

Computation: Scores: Sample I : 3, 6, 6, 7, 8, 8, 8, 10, 12, 13, 13, 16, 16, 18, 20

Sample II : 6, 8, 9, 13, 13, 14, 15, 16, 16, 18, 18, 19, 19, 24, 26.

Combined median = $(14 + 15)/2 = 14.5$

	Group I	Group II	
No. above combined Mdn.	(A) 6	(B) 9	15 A+B
No. below Combined Mdn.	(C) 11	(D) 4	15 C+D
	17 A+C	13 B+D	N=30

$$\begin{aligned}
 \text{Formula: Chi square} &= \frac{N(|AD - BC| - N/2)^2}{(A+B)(C+D)(A+C)(B+D)} \\
 &= \frac{30(|6 \times 4 - 9 \times 11| - 30/2)^2}{15 \times 15 \times 17 \times 13} \\
 &= 2.3
 \end{aligned}$$

Interpretation: Since the observed Chi square of 2.3 is less than 3.84, accept H_0 . The two samples do not differ significantly with respect to their medians.

10.9.4. A General New Parametric Test for Two Independent Samples—Run Test

The Sign Test and the Median Test were introduced as non-parametric tests for location in which two samples were compared in terms of central tendency to test whether they were drawn from the same population.

The Run Test to be described below is a general two sample test for testing the null hypothesis that two independent samples come from identical populations against the alternative hypothesis that the two populations differ in any manner whatsoever i.e., in central tendency, in dispersion, in skewness, in kurtosis or in any other way. The Run Test is less powerful in disclosing differences of a particular kind, for it is a test of any sort of differences and not for a particular type of difference. Hence, if we are interested in testing whether the two populations differ only in one particular respect, say either in central tendency or in dispersion, then we should use a test of location or dispersion from amongst the tests described earlier.

In 50 tosses of a coin, the following distribution of Heads (H's) and Tails (T's) was obtained. Use a Run Test to see whether the coin was a biased one which produced a non-random distribution of H's and T's.

HHH TT H TTTTT HHH T H TTTT HHH TTTT H TT
HH TTTT HHH TTTT H TT HH T H

Steps

1. The observations as obtained have been listed above as H's and T's. A run is an observation in the sequence of

letters* of the same kind which cannot be extended by incorporating an adjacent observation. In the example above, the H runs have been underlined and T runs overscored. If the two samples are from a common population, the H's and T's will generally be well mixed and the number of runs will be large. Now count the number of runs and call them as v .

Number of runs for H's = 11; and for T's = 10.

Total No. of runs = $v = 11 + 10 = 21$

$N_1 = 21$; $N_2 = 29$ (N_1 and N_2 are number of H's and T's respectively)

2. *Hypothesis*: H_0 : The coin is not a biased one
 H_1 : The coin is a biased one.
3. *Computation* : If $N_1 > 10$ and $N_2 > 10$, the test of significance is obtained by taking v to be normally distributed with Mean

$$\mu_v = \frac{2N_1N_2}{N_1 + N_2} + 1 \quad (10.10)$$

and variance,

$$\sigma_v^2 = \frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1 + N_2)^2(N_1 + N_2 - 1)} \quad (10.11)$$

In the example given :

$$\mu_v = \frac{2(21 \times 29)}{21 + 29} + 1 = 25.36$$

$$\sigma_v^2 = \frac{2 \times 21 \times 29(2 \times 21 \times 29 - 21 - 29)}{(21 + 29)^2(21 + 29 - 1)}$$

$$= 11.61$$

$$\sigma_v = \sqrt{\sigma_v^2} = \sqrt{11.61} = 3.41$$

$$z = \frac{v - \mu_v}{\sigma_v} = \frac{21 - 25.36}{3.41} = \frac{4.36}{3.41} = 1.28$$

4. *Interpretation:* The calculated value of z , (1.28), is less than the critical value of z required for significance at .05 level (critical $z=1.96$) hence the null hypothesis cannot be rejected.

Therefore the coin is not biased and has produced a random distribution of H's and T's.

A correction for continuity as given below may be used when $N_1 + N_2$ is not very large.

$$z = \frac{|v - \mu_v| - .5}{\sigma_v} \quad (10.12)$$

If scores of two samples are given, the scores can be arranged in an ascending order in a common arrangement keeping the identity of the score as to which group it belonged. Then the runs for group I and group II can be counted and the procedure given above repeated. For example:

Scores

Group A : 5, 6, 7, 8, 12, 12, 12, 13, 18, 19
 Group B : 2, 3, 4, 10, 10, 10, 11, 11, 14, 15

The arrangement for counting of runs will be as follows:

Group	<u>B B B</u>	<u>A A A A</u>	<u>B B B B B</u>	<u>A A A A</u>
Score	2, 3, 4	5, 6, 7, 8	10, 10, 10, 11, 11	12, 12, 12, 13

Group	<u>B</u>	<u>B</u>	<u>A</u>	<u>A</u>
Score	14,	15	18,	19

Number of runs for Group A=3, and for B=4

Total No. of runs or $v=3+4=7$

Repeat the other steps.

10.9.5. The Kolmogorov—Smirnov Two Sample Test

The Kolmogorov-Smirnov (K—S) Test is quite useful in applications in which two distributions compared are both samples. The feature primarily compared is the numerical level on some scale. Differences in variance and kurtosis have little effect on the test. Ordered categories are sufficient refinement and no assumption about equal intervals is required. The

null hypothesis is that the two distributions arose by random sampling from the same population. The test differs from small to large size samples; and for one-tailed to two-tailed situations.

The K-S Test with Small Samples

When N is 40 or less in each of the two distributions, the K-S test is applied as a small sample test. It is more convenient when $N_1 = N_2$ because tables of critical K statistic exist to facilitate its use.

Example: Two groups of boys and girls with $n=12$ in each case had a cricket match and scored the runs presented in two frequency distribution in Table 10.4. Have the two distributions arisen by random sampling from the same population ?

TABLE 10.4

The Kolmogorov-Smirnov Test of Similarity of Distributions (Two independent small samples with equal n 's)

Score: No. of runs made	Frequencies (f)		Cum. f .		K_c
	Boys	Girls	Boys	Girls	
55-59	2	0	12	12	0
50-54	1	0	10	12	2
45-49	3	0	9	12	3
40-44	2	1	6	12	6
35-39	1	3	4	11	7
30-34	2	2	3	8	5
25-29	1	4	1	6	5
20-24	0	2	0	2	2

$K=7$

Hypotheses: H_0 : There is no significant difference in the two distributions of runs scored by boys and girls.

H_1 : There is a significant difference in the two distributions of runs scored by boys and girls.

It is a two tailed situation.

Computation: The calculations are quite easy. The essential step is to find out cumulative frequency distributions for the two samples, as shown in Table 10.4. The last operation is to find the category differences or K_C . The largest K_C is statistic K . In this case $K=7$.

Interpretation

Consulting Table J with $N_1=N_2=10$, we find that a $K=7$ is required for significance at .05 level in a two-tailed situation. A $K=8$ is required at .01 level.

Since the calculated value of $K=7$, the Null hypothesis is rejected at .05 level. However, the calculated K is not significant at .01 level as it is less than 8.

One-tailed situation:

Suppose our hypothesis was that the differences were in favour of boys, we would have used a one-tailed test which would require $K=6$ at .05 level and $K=7$ at .01 level to be significant. Since the calculated $K=7$, we would have rejected H_0 at .01 level, which is higher than the level of rejection in a two tailed situation.

10.9.6 The K-S Test with Large Samples

When both the groups have n 's larger than 40, a large sample $K-S$ test is warranted. An example with unequal n 's has been chosen to illustrate the generality of the procedure. Both a one-tail and two-tail tests have been applied. The test uses Chi-square distribution.

Example: A test of extroversion was administered to a set of 40 students identified by their teachers as front benchers in

the class and to 50 students identified as back-benchers. The scores so obtained have been shown in Table 10.5 below in the form of two frequency distributions. Do the two groups belong to the same population ?

TABLE 10.5

**Kolmogorov-Smirnov Two-Sample Test with
Large and Unequal N's**

Score*	Frequency of		Cf_A	Cf_B	Cp_A	Cp_B	d_c
	Gr. A*	Gr. B*					
19	3	0	40	50	1.000	1.000	.000
18	5	0	37	50	.925	1.000	.075
17	6	0	32	50	.800	1.000	.200
16	10	5	26	50	.650	1.000	.350
15	8	4	16	45	.400	.900	.500
14	3	8	8	41	.200	.820	.620
13	2	10	5	33	.125	.660	.535
12	1	13	3	23	.075	.460	.385
11	2	10	2	10	.050	.200	.150
10	0	0	0	0	.000	.000	.000
	40	50					D=.620

*A=Front benchers; B=Back benchers; Scores: Extroversion scores

Hypothesis: H_0 : The two groups of front benchers and back benchers belong to a common population.

H_1 : The two groups of front benchers and back benchers do not belong to a common population. These are non-directional two-tail situations.

Computation: The major step is to obtain cum. frequencies and convert them into cum. proportions for each of

the two groups. Cum. proportions can be obtained by dividing each category cum. f by the n of its groups. Then the d_c values or the category differences are obtained by subtracting category cum. proportions. The largest d_c is taken as D statistic. In our case, $D = .620$.

Calculations of critical values of D in two tailed test are obtained by using the following formulas:

<i>Sig. level</i>	<i>Critical D value</i>	
.10	$1.22 \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$	(10.13)

.05	$1.36 \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$	(10.14)
-----	---	---------

.01	$1.63 \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$	(10.15)
-----	---	---------

.001	$1.95 \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$	(10.16)
------	---	---------

The radical term may be solved to facilitate the calculation of critical values for various levels. In our case, $N_1 = 40$, and $N_2 = 50$, the radical value,

$$\sqrt{\frac{N_1 + N_2}{N_1 N_2}} = \sqrt{\frac{40 + 50}{40 \times 50}} = .212$$

Hence, critical values of D can be obtained by multiplying the radical with the numerical values given in the formulas. If we choose .05 level for interpretation, we obtain

$$D = 1.36 \times .212 = .288$$

Interpretation: The obtained value of D (.520) is larger than the critical value of D (.288); hence the H_0 cannot be retained. The two groups differ significantly on extroversion and therefore do not come from the common population.

One-tail case

If the hypothesis was that the front benchers are likely to be more extroverted than the back benchers, a one-tail test would have been warranted. For this purpose a Chi-square can be derived from D by means of the formula:

$$\chi^2 = 4D^2 \left(\frac{N_1 N_2}{N_1 + N_2} \right) \quad (10.17)$$

where N_1 and N_2 are number of cases in the two groups; and D is the largest difference of category proportions. In our example

$$\begin{aligned} \chi^2 &= 4(.620)^2 \left(\frac{40 \times 50}{40 + 50} \right) \\ &= 34.17 \end{aligned}$$

With $df=2$, the calculated Chi square is much greater than the critical value of Chi square required for significance at .01 level. Hence it is significant. The front benchers are significantly higher than the back benchers on extroversion and the two groups do not come from a common population.

10.9.7 Some Precautions

In the use of the non-parametric tests, the student is cautioned against the following lapses:

1. When measurements are in terms of interval and ratio scales, the transformation of the measurements on nominal or ordinal scales will lead to the loss of much of information. Hence, as far as possible parametric tests should be applied in such situations. In using a non-parametric method as a shortcut, we are throwing away dollars in order to save pennies.
2. In situations where the assumptions underlying a parametric test are satisfied and both parametric and non-parametric tests can be applied, the choice should fall on the parametric test because most parametric tests have greater power in such situations.

3. Non-parametric tests, no doubt, provide a means for avoiding the assumption of normality of distribution. But these methods do nothing to avoid the assumptions of independence on homoscedasticity wherever applicable.
4. The behavioural scientist should specify the null hypothesis, alternative hypothesis, statistical test, sampling distribution, and level of significance in advance of the collection of data. Hunting around for a statistical test after the data have been collected tends to maximise the effects of any chance differences which favour one test over another. As a result, the possibility of rejecting the null hypothesis when it is true (Type I error) is greatly increased. However, this caution is applicable equally to parametric as well as non-parametric tests.
5. We do not have the problem of choosing statistical tests for categorical variables. Non-parametric tests alone are suitable for enumerative data.
6. The F and t tests are generally considered to be robust tests because the violation of the underlying assumptions does not invalidate the inferences. It is customary to justify the use of a normal theory test in a situation where normality cannot be guaranteed, by arguing that it is robust under non-normality.

10.9.8 A Guide for the Selection of Non-Parametric Tests

In this Chapter, only a few more popular non-parametric tests have been used. However, a general guide for the selection of non-parametric tests is presented in Table 10.6. It is based on the level of measurement, single or multiple samples, and correlated or independent samples. Several volumes are available in which these tests have been described.

TABLE 10.6
A General Guide for the Selection of Non-Parametric Tests

Level of Measure- ment	Non-Parametric Statistical Test*					Non-parametric measure of correlation
	One-sample case	Two-Sample case		K-Sample case		
		Related samples	Independent samples	Related samples	Independent samples	
Nominal	Binomial test, χ^2 one-sample test	McNemar test for the signi- ficance of changes	Fisher exact probability test; χ^2 test for two in- dependent samples	Cochran Q test	χ^2 test for k independent samples	Contingency coefficient: C
Ordinal	Kolmogrov- Smirnov one-sample test, one-	Sign test, Wilcoxon matched pairs	Median test, Mann- Whitney U test, Kolmo-	Friedman two-way analysis of	Extension of the median test, Kruskal- Wallis	Spearman rank correlation coefficient: r, Kendall rank correlation coefficient: r

Interval	sample runs test	signed-ranks test	grov-Smirnov two-sample test, Wald-Wolfowitz runs test, Moses test of extreme reactions	variance	one-way analysis of variance,	Kendall partial rank correlation coefficient: r_{xy-z} Kendall coefficient of concordance: W.
	Walsh test, Randomization test for matched pairs	Randomization test for two independent samples				

*Each column lists, cumulatively downward, the tests applicable to the given level of measurement. For example, in the case of k related sample, when ordinal measurement has been achieved both the Friedman two-way analysis of variance and the Cochran Q test are applicable. (Siegel, 1956)

Exercises for Practice

- 10.1 Seventy-nine urban and 83 rural college students were asked to respond on a three-point scale—Approve, Neutral and Disapprove—to a question “Should sex education be included in the college courses.” Their number in each category is shown below:

	Urban	Rural
Approve	58	35
Neutral	11	25
Disapprove	10	23

Is the pattern of response independent of the residence of the students?

- 10.2 Eighty-two teachers were classified into three categories—very good, average, and poor—by a consensus of team of inspectors. If the teaching proficiency is distributed normally, does, this distribution differ significantly from the normal distribution? The observed frequencies are:

Good=32; Average=40; Poor=12.

- 10.3 Two hundred students were asked whether open book examinations be instituted at the Post-graduate stage. Their responses are given below:

Strongly approve	Approve	Indifferent	Disapprove	Strongly disapprove
46	36	48	34	36

Do these results differ significantly from equal preferences in the group?

- 10.4 The following are observations for two independent samples:

Sample I: 5, 6, 8, 10, 14, 15, 19, 22, 25, 25

Sample II: 8, 10, 12, 16, 18, 18, 27

Use Median test to find out if the two samples came from the same population with respect to their medians.

- 10.5 The number of attempts taken by members of two groups of boys and girls with 20 persons in each group, in hitting a shooting target is given below. Find out if the two groups differed significantly in their shooting ability (or the boys are more efficient than the girls)

Number of attempts		
Persons	Boys	Girls
1	1	8
2	2	8
3	3	8
4	3	9
5	4	10
6	5	11
7	7	12
8	7	12
9	7	12
10	7	14
11	8	14
12	8	14
13	8	16
14	8	16
15	9	16
16	14	18
17	14	18
18	14	20
19	15	20
20	15	22

- 10.6 On the last day of depositing the college fee for the month of January 1985, a large queue of boys and girls was seen at the fee counter. Their sex was recorded according to their position in the queue. Use a Run Test to see whether the arrangement of their position differed significantly from randomness.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13
Sex	B	B	G	B	G	B	B	G	G	G	G	B	B
Position	14	15	16	17	18	19	20	21	22	23	24	25	
Sex	B	G	B	G	G	B	B	B	B	G	G	G	

- 10.7 Two groups of boys and girls were administered a test of attitude towards sex in movies. The higher score means greater preference for sex. Distributions of their scores are given below. Do the two groups belong to a common population with regard to their attitude towards sex in movies?

Attitude score	10	11	12	13	14	15	16	17	18	19	20
Boys	1	5	6	8	7	10	2	1	0	0	0
Girls	0	0	0	2	3	5	8	10	1	1	0

Use One-tail and Two-tail tests.

CHAPTER 11

THE ANALYSIS OF VARIANCE, ANOVA

11.1 The Rationale

The *t* test of significance is adequate for any experiment that involves only two groups and only a single factor. It provides only a test of a single mean difference. But suppose we have an experimental design involving three groups, A, B and C, with each group tested after a different experimental treatment or under a different set of experimental conditions. The use of *t* test as a relatively simple statistical technique would still be possible and would involve taking two group means at a time and testing the significance of the difference. The number of mean comparisons in this case would be three viz., A and B; A and C; and B and C. However, the problem arises when the number of groups is larger say five or more. If we have ten groups, the number of comparisons which would be required under the *t* test would be given by the formula:

$$\frac{N(N-1)}{2} = \frac{10 \times 9}{2} = 45.$$

Obviously, some method of testing differences among all of the means at the same time would prove very valuable. The analysis of variance and the corresponding test of significance based upon *F* distribution permit us to do this. Another important consideration which rules out the use of *t* test and warrants the use of the more sophisticated *F* test of significance, is the case of situations in which two or more experimental variables or one experimental and one or more control variables are simultaneously operating and not only comparison of means within each variable is required but also the joint operation or interaction of two or more variables is of interest.

The technique of analysis of variance was first devised by Sir Ronald Fisher, an English statistician who is also considered to be the father of modern statistics as applied to social and behavioural sciences. It was first reported in 1923 and its early applications were in the field of agriculture. Since then it has found wide applications in many areas of experimentation. The analysis of variance, as the name indicates, deals with variances rather than with standard deviations and standard errors. It is a method of dividing the variation observed in experimental data into different parts, each part assignable to a known source, cause, or factor. We may assess the relative magnitude of variation resulting from different sources and ascertain whether a particular part of the variation was greater than expectation under the null hypothesis. However, it must be remembered that analysis of variance divides the total sum of squares, $\sum (X - \bar{x})^2$, into additive parts which are then converted into *mean square* simply by dividing the sum of squares by the relevant degrees of freedom. The main difference between variance and mean square is that the former is obtained by dividing the sum of squares by n while the latter is obtained by dividing the sum of squares by degrees of freedom. Hence, it is advisable to keep this fact in view while understanding the rationale of the analysis of variance.

In its simplest form, the analysis of variance is used to test the significance of the differences among means of a number of different groups supposed to have come from different populations. The total sum of squares is analysed into two parts: a sum of squares based upon variation *within the several groups*; and a sum of squares based upon variation *between the group means*. Then from the two sums of squares, independent estimates of the population variance are computed. The value of F is, then, the ratio between the two estimates of the population variance

$$F = \frac{\sigma^2_{\text{Between groups}}}{\sigma^2_{\text{Within groups}}} \quad \text{or} \quad \frac{\sigma^2_B}{\sigma^2_W} \quad (11.1)$$

in which σ^2 is the standard symbol of population variance.

11.2 One Way or Single Classification ANOVA

In its simplest form ANOVA can be used when a number of treatments based on a single factor are involved. For example, in a field experiment, three randomly selected groups have been assigned randomly to three different experimental treatments say, traditional method; programmed learning method; and multi-media method. At the end, the criterion scores are obtained. The mean scores of three groups can then be compared by using ANOVA. Since only one factor, i.e., method of teaching (of course with three variations of the method) is involved, the situation warrants a single classification or one way ANOVA and can be diagrammed as below:

<i>Programmed Learning Method</i>	<i>Multi-media Method</i>	<i>Traditional Method</i>
X_{11}	X_{12}	X_{13}
X_{21}	X_{22}	X_{23}
X_{31}	X_{32}	X_{33}
\vdots		\vdots
$X_{n_1,1}$	$X_{n_2,2}$	$X_{n_3,3}$

In which X stands for scores; subscripts, for individuals and columns; and n_1, n_2, n_3 , for number of persons in each group.

As stated earlier, in the simplest case, the total variance is partitioned into two components: *The between variance* which is the variance of the means of the experimental groups about the grand mean of the total sample; this component includes the contribution of the various experimental treatments plus variance due to sampling fluctuations or error variance. The other component, *the within variance* is an average of the variance within the experimental groups. It is often termed as *error variance* also because it results from sampling fluctuations i.e., due only to individual differences. It may be expressed as σ_e^2 or variance due to random fluctuations in sampling.

However, in the formula for F-ratio, σ_w^2 instead of σ_e^2 is generally used.

The between variance can be expressed as:

$$\sigma_B^2 = \sigma_e^2 + \sigma_{Tr}^2 \quad (11.2)$$

in which σ_{Tr}^2 stands for variance due to experimental treatment; σ_e^2 , for error variance.

The value of F is simply the ratio between the two variances—between and within. This ratio generates a new sampling distribution of F . The test whether the experimental treatment caused significantly different results is made by comparing $\sigma_B^2 / \sigma_W^2 = (\sigma_e^2 + \sigma_{Tr}^2) / \sigma_e^2$ to the F distribution with appropriate degrees of freedom for numerator and denominator. If the scores are truly scores from the same population, the two variance estimates, the σ_B^2 and σ_W^2 , will be the same. However, from actual experimental data, it is difficult to find samples, though randomly and independently chosen, having exactly the same means and variances. Therefore, random differences between the numerator and denominator of the F ratio are to be expected and are given in the tables of F -distribution. If the treatments of the experiment produce decided differences in variance among criterion measures, then the between variance will be increased leading to the increase in the value of F -ratio which would then be greater than one. A comparison of this F -ratio with the appropriate value in the F -distribution table can establish whether the former arose due to chance or due to real differences among the various treatments. With an initial arbitrary decision concerning the acceptable level of the probability, one can decide to reject or not to reject the null hypothesis.

In a one-way ANOVA situation, the total variance is equal to the variance within groups and variance between groups. This concept will be demonstrated through a numerical example given in Table 11.1. There are two procedures of calculating total, between, and within variances—one through

a raw score approach and the other through a deviation score approach. Both these approaches will be demonstrated below:

11.2.1 Deviation Score Method

TABLE 11.1

**Work-sheet for One-Way Analysis of Variance on
Hypothetical Scores using Deviation
Score Method**

<i>Step I The Measurements (X)</i>			
	<i>Set I</i>	<i>Set II</i>	<i>Set III</i>
	10	3	10
	7	3	11
	6	3	10
	10	3	5
	4	3	6
	3	3	8
	2	3	9
	1	3	12
	8	3	9
	9	3	10
ΣX_s	60	30	90 $\Sigma X = 180$
M_s	6	3	9 $M_t = 6$
<i>Step II Deviations within sets (x_s)</i>			
	+4	0	+1
	+1	0	+2
	0	0	+1
	+4	0	-4
	-2	0	-3
	-3	0	-1
	-4	0	0
	-5	0	+3
	+2	0	0
	+3	0	+1

Step III Squares of deviations within sets (x_s^2)

16	0	1	
1	0	4	
0	0	1	
16	0	16	
4	0	9	
9	0	1	
16	0	0	
25	0	9	
4	0	0	
9	0	1	
<hr/>			
100	0	42	$\Sigma x_s^2 = 142$

Step IV Deviations of set means from grand mean (d)

d	0	3	-3	
d^2	0	9	9	$\Sigma d^2 = 18$
nd^2	0	90	90	$n \Sigma d^2 = 180$

In Table 11.1, fictitious data have been given. In set II, all values have been kept equal. The purpose is to explain an important point later in the discussion. The three sets of scores were hypothetically obtained under three different conditions or treatments. The hypothesis to be tested is whether all the observations came by random sampling from the same general population or were there systematic overall differences among the three set means. In symbolic terms:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu$$

In verbal terms, the null hypothesis states that the means of the three populations do not differ among themselves and are equal to the mean of the general population. The alternative hypothesis expresses that these means are not equal.

Steps of Computation (Deviation Score Method)**A. Calculation of Sum of Squares Within Sets:**

- I. Compute the sums and means of all the three sets, the grand total ΣX , and the grand mean M_r .
- II. For every set, compute deviations from the set mean M_x by using the formula $x = (X - M_x)$. Designate them as x_i or deviations within sets.
- III. Square the deviations within sets to find each x_i^2 . Add them to obtain Σx_i^2 . These are sum of squares of deviations within sets.

B. Calculation of Sum of Squares Between Sets:

- IV. For each set, take the deviation between each set mean and grand mean, $(M_s - M_r)$, and call them d . Square each d and sum them up to obtain Σd^2 . Multiply each d^2 by n , the number of scores in each set and sum them up to obtain $n\Sigma d^2$. This value is sum of squares between sets.

C. For the Calculation of Total Sum of Squares

The total sum of squares can be obtained by adding the sum of squares within sets and the sum of squares between sets. However, for the purpose of verification of the relationship among these three types of sums of squares, the total sum of squares can be calculated directly by subtracting each score from the grand mean, squaring the deviation, and summing all the deviations up. In our example,

$$\begin{aligned}\text{The total sum of squares} &= (10-6)^2 + (7-6)^2 + \dots + (10-6)^2 \\ &= 322.00\end{aligned}$$

It checks with the value obtained by adding sum of squares within sets, and sum of squares between sets. The student should thus understand the following relationships:

$$(i) \quad \text{Total sum of squares} = \text{Within sum of squares} + \text{Between sum of squares}$$

$$\text{or } SS_T = SS_W + SS_B \quad \dots(11.3)$$

$$(ii) \quad \text{Between sets sum of squares} = \text{Total sum of squares} - \text{Within sum of squares}$$

$$\text{or } SS_B = SS_T - SS_W \quad \dots(11.4)$$

$$(iii) \quad \text{Within sets sum of squares} = \text{Total sum of squares} - \text{Between sum of squares}$$

$$\text{or } SS_W = SS_T - SS_B \quad \dots(11.5).$$

D. Calculation of Degrees of Freedom and Mean Square:

Degrees of freedom can be calculated as follows:

$$df_{total} = (\text{Total number of scores} - 1); \text{ or } (N - 1); \text{ or } (30 - 1) = 29$$

$$df_{between} = (\text{No. of sets} - 1); \text{ or } (s - 1); \text{ or } (3 - 1) = 2$$

$$df_{within} = (\text{number of scores in each set minus 1, multiplied by number of sets}); s(n - 1); 3(10 - 1) = 27$$

The sum of degrees of freedom for between and within sets should add up to the total degrees of freedom. The mean squares for each source of variation can be obtained by dividing the sum of squares by their respective df.

With all the above computational results at our command, we can set up a table of summary of ANOVA results as given below:

TABLE 11.2

Summary of Analysis of Variance

Component (Sources of variation)	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Between sets	180.00	2	90.00	17.11
Within sets	142.00	27	5.26	
Total	322.00	29		

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{90.00}{5.26} = 17.11; \text{ df } = 2, 27$$

Interpretation: The calculated F value of 17.11 is to be compared with the value of F required for significance at .05 and .01 level. For this purpose, Table K in appendix is to be consulted. Reading of any F value from the table depends upon the df for greater mean square i.e., between sets; and df for smaller mean square i.e., the within mean square. The former are given vertically in the columns, and the latter horizontally in the rows. The value at the intersection of the two is the value of F at a particular level of significance with df as mentioned above. In Table K, the upper values are at .05 level, while the lower ones are at .01 level. In our case, the values of F from the table are $_{.05}F_{(2,27)} = 3.35$; $_{.01}F_{(2,27)} = 5.49$.

The calculated value of F is higher than both the values of F at .05 and .01 levels. Since .01 level is higher, we can interpret that the calculated value of F is significant at .01 level. Hence the H_0 cannot be accepted. It can be said that the overall differences among the three set means are significant and not due to chance. The three sets do not belong to the same population with regard to their means.

As pointed out earlier, all scores in set II were purposely kept equal. The main objective was to explain an important

point. This set has a zero within variance because all scores are equal to the set mean and hence it does not contribute anything to the variance within sets. In the same manner, the mean of set I is equal to the grand mean, and hence does not contribute anything to the variance between sets.

11.2.2 Raw Score Method

TABLE 11.2 A

Worksheet for One-way ANOVA on Hypothetical Scores (RAW SCORE METHOD)

<i>Set I</i>		<i>Set II</i>		<i>Set III</i>		
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	
10	100	3	9	10	100	
7	49	3	9	11	121	
6	36	3	9	10	100	
10	100	3	9	5	25	
4	16	3	9	6	36	
3	9	3	9	8	64	
2	4	3	9	9	81	
1	1	3	9	12	144	
8	64	3	9	9	81	
9	81	3	9	10	100	
Σ 's	60	460	30	90	90	852
	(ΣX_1)	(ΣX_1^2)	(ΣX_2)	(ΣX_2^2)	(ΣX_3)	(ΣX_3^2)
M 's	6		3		9	

$N=30$; $n_1=n_2=n_3=10$: ΣX or $T=60+30+90=180$.

A. Sum of Squares

$$\begin{aligned}
 1. \text{ Correction term, } C &= \frac{(\Sigma X)^2}{N} = \frac{T^2}{N} \quad \dots(11.6) \\
 &= \frac{(180)^2}{30} = 1080
 \end{aligned}$$

$$2. \text{ Total sum of Squares, } SS_T = \Sigma X^2 - C \quad \dots(11.7)$$

$$\therefore SS_T = 10^2 + 7^2 + 6^2 + \dots + 12^2 + 9^2 + 10^2 - 1080 \\ = 1402 - 1080 = 322.$$

$$3. \text{ Sum of squares among set means,}$$

$$SS_B = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - C \quad \dots(11.8) \\ = \frac{(60)^2}{10} + \frac{(30)^2}{10} + \frac{(90)^2}{10} - 1080 \\ = 1260 - 1080 = 180.00$$

$$4. \text{ Sum of squares within sets,}$$

$$SS_W = SS_T - SS_B \\ = 322 - 180 \\ = 142$$

B. Summary of Analysis of Variance

Source of variation	df	Sum of squares SS	Mean square (Variance)
Between sets	2	180	90
Within sets	27	142	5.26
Total	29	322	

$$F = \frac{90}{5.26} = 17.11$$

Steps for computation (Raw Score Method)

- I. Set up the scores under X_1 , X_2 and X_3 as shown in Table 11.2.
- II. Square the scores of all the sets and write under X_1^2 , X_2^2 , and X_3^2 .
- III. Obtain all sums by adding up the individual columns of the table. Obtain grand sum of scores, Σx , T also.
- IV. After these intermediate calculations, we are now ready for the calculation of various sums of squares.

- V. Calculate correction term, C as it is required for the adjustment of origin when raw scores are used for the calculation of variance. In fact, the assumed mean here is zero, and all scores form deviations from zero.
- VI. Calculate SS_T which is sum of squared scores minus correction.
- VII. Square each set sum and divide by respective n , and over all the three sets. Deduct C , to obtain SS_B . In case of equal n 's a common denominator, n , can be used as below:

$$SS_B = \frac{(\sum X_1)^2 + (\sum X_2)^2 + \dots + (\sum X_k)^2}{n} - C \quad (11.9)$$

- VIII. Subtract SS_B from SS_T to obtain SS_W . SS_W can also be calculated directly without computing SS_T .

$$SS_W = \frac{\sum X^2 - (\sum X_1)^2 + (\sum X_2)^2 + \dots + (\sum X_k)^2}{n} \quad (11.10)$$

- IX. Table of summary of ANOVA can be set up as explained earlier in the case of deviation score method. Degrees of freedom, MS and F can also be calculated similarly.

All sums of squares, mean square and the value of F are the same as obtained by using the deviation score method. The raw score method is to be preferred when score values are small, the means are not whole numbers, and a calculating machine is to be used. However, with small data and means as whole numbers, the deviation score method can be used with profit.

11.3 Post-Anova Test of Differences by Use of 't'

After a significant F has been obtained, one can look for significant pairs of means. For this purpose, t test can be used. Since MS_W , the best estimate of the population variance, is available from the ANOVA summary table, the standard deviation can be readily computed as below:

$$\sigma_W = \sqrt{MS_W} \quad (11.11)$$

Standard error of difference of means is given by,

$$SE_D = \sigma_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (11.12)$$

$$= 2.29 \sqrt{\frac{1}{10} + \frac{1}{10}} = 1.024$$

Value of t with $df=27$ (Related with MS_w)

at .05 level = 2.05; at .01 level = 2.77

Now *critical mean difference* can be calculated for comparison with all mean differences to find out the significant pairs of means.

$$D_{.05} = t_{.05} \times SE_D \quad (11.13)$$

$$= 2.05 \times 1.024 = 2.099$$

$$D_{.01} = t_{.01} \times SE_D \quad (11.14)$$

$$= 2.77 \times 1.024 = 2.836$$

Now set up a table of mean differences as below:

Sets	Mean Differences		
	I	II	III
Means	6.0	3.0	9.0
6.0		3.0*	3.0*
3.0		—	6.0*
9.0			—

*Significant at .01 level.

All the mean differences are larger than the critical mean difference, 2.836 (at .01 level). Hence, all the mean differences are significant at .01 level. When number of sets is large, the above procedure saves time and shows, at a glance, the pairs of means which are significant.

11.4 Two-way or Double Classification Anova

In experiments more than one experimental factors or one experimental and one or more control factors may be used. For example, in a field experiment, three different methods of teaching—programmed learning, multi-media, and traditional—may be tried out as an experimental factor. At the same time, two different teachers (teacher factor), may be used as a control factor. Thus, we have three levels of method-factor and two levels of teacher-factor and there are in all $3 \times 2 = 6$ combinations of these two factors. This situation is termed as a 3×2 factorial design. Depending upon the number of factors and the number of levels of each factor, several variations of the factorial designs are possible. In this section, only two-way or double classification analysis will be presented. The hypothetical data used in one-way classification ANOVA in Table 11.1 has been once again used by introducing a second factor of teacher. The three sets of scores have been designated as programmed learning (M_1), multi-media (M_2), and traditional (M_3) and the two levels of the teacher-factor as Teacher I (T_1) and Teacher II (T_2).

The factorial design will thus be as follows:

		Method		
		M_1	M_2	M_3
Teacher	T_1	M_1T_1	M_2T_1	M_3T_1
	T_2	M_1T_2	M_2T_2	M_3T_2

The mathematical model for this example will be:

$$X = \mu + d_m + d_t + d_{mt} + e_r \quad (11.15)$$

In which symbols are:

X = Any raw score; μ = Population mean or grand mean;
 d_m = deviation due to method factor; d_t = deviation due to teacher factor; d_{mt} = deviation due to combined effect of m and t or interaction of m and t ; and e_r = random error.

In verbal expression, any raw score is a combination of the population mean (here grand mean), plus variation due to method factor plus variation due to teacher factor, plus variation due to the combined effect of method and teacher factors plus random error due to sampling fluctuations. This model can be further extended to three factor or larger factorial designs. The model makes it obvious that the various effects are additive and through a process of analysis, variance due to various effects can be separately determined and tested for significance. The population mean, μ is the grandmean of the scores empirically obtained, and is a constant value and thus does not contribute to any variation in the data.

The procedure of calculation of various effects in the two-way ANOVA is illustrated in Table 11.3. The notation for the purpose is also given therein.

TABLE 11.3

Worksheet for two-way Analysis of Variance on Hypothetical Data (RAW SCORE METHOD)

	Teaching Methods					Total	
	Programmed Learning (M_1)		Multi-media (M_2)		Traditional (M_3)		
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	
Teacher I (T_1)	10	100	3	9	10	100	
	7	49	3	9	11	121	
	6	36	3	9	10	100	
	10	100	3	9	5	25	
	4	16	3	9	6	36	
Σs	37	301	15	45	42	382	94
Teacher II (T_2)	3	9	3	9	8	64	
	2	4	3	9	9	81	
	1	1	3	9	12	144	
	8	64	3	9	9	81	
	9	81	3	9	10	100	
Σs	23	159	15	45	48	470	86
Σ	60	460	30	90	90	852	180

Total No. of observations, $N=30$

No. of observations in each cell, $n=5$

No. of columns (teaching methods), $m=3$

No. of rows (teachers), $t=2$.

The student should be able to calculate and identify various sums of scores and sums of squared scores which are required in the further calculations of sums of squares:

Sums for Method Factor

$$\Sigma X_{M_1} = 60; \Sigma X_{M_2} = 30; \Sigma X_{M_3} = 90: \text{Total } 180$$

Sums for Teacher Factor

$$\Sigma X_{T_1} = 94; \Sigma X_{T_2} = 86: \text{Total } 180$$

Sums for Cells (Method \times Teacher)

$$\Sigma X_{M_1 T_1} = 37; \Sigma X_{M_1 T_2} = 23; \Sigma X_{M_2 T_1} = 15; \Sigma X_{M_2 T_2} = 15;$$

$$\Sigma X_{M_3 T_1} = 42; \Sigma X_{M_3 T_2} = 48: \text{Total} = 180$$

Sums of Squared Scores

$$\Sigma X_1^2 + \Sigma X_2^2 + \dots + \Sigma X_N^2 = 10^2 + 7^2 + \dots + 10^2 = 1402$$

A. *Sum of Squares*

1. Correction term,

$$C = \frac{(\Sigma X)^2}{N} = \frac{T^2}{N} = \frac{(180)^2}{30} = 1080.00$$

2. Total sum of squares $= \Sigma X^2 - C$

$$\begin{aligned} SS_{\text{total}} &= (10^2 + 7^2 + 6^2 + 10^2 + \dots + 9^2 + 12^2 + 9^2 + 10^2) \\ &= 1080 \end{aligned}$$

$$= 1402 - 1080 = 322.00$$

3. Sum of the Squares between cells,

$$SS_{\text{teacher} \cdot \text{method}} = \frac{\Sigma (\Sigma X_{mt})^2}{n} - C \quad (11.16)$$

$$\begin{aligned}
 &= \frac{(37)^2 + (15)^2 + (42)^2 + (23)^2 + (15)^2 + (48)^2}{5} - 1080 \\
 &= \frac{1369 + 225 + 1764 + 529 + 225 + 2304}{5} - 1080 \\
 &= \frac{6416}{5} - 1080 \\
 &= 1283.2 - 1080 \\
 &= 203.2
 \end{aligned}$$

4. Sum of squares between rows (teacher)

$$\begin{aligned}
 SS_{\text{teacher}} &= \frac{\sum (\sum X_i)^2}{nm} - C & (11.17) \\
 &= \frac{(94)^2 + (86)^2}{5 \times 3} - 1080 \\
 &= 1082.133 - 1080 = 2.133
 \end{aligned}$$

5. Sum of Squares between columns (Method)

$$\begin{aligned}
 SS_{\text{method}} &= \frac{\sum (\sum X_m)^2}{nt} - C \\
 &= \frac{(60)^2 + (30)^2 + (90)^2}{5 \times 2} - 1080 \\
 &= 1260 - 1080 \\
 &= 180
 \end{aligned}$$

6. Sum of Squares for interaction

$$\begin{aligned}
 SS_{\text{teacher} \times \text{method}} &= SS_{\text{teacher} \cdot \text{method}} - SS_{\text{teacher}} - SS_{\text{method}} \\
 &= 203.2 - 2.133 - 180 \\
 &= 203.2 - 182.133 \\
 &= 21.067
 \end{aligned}$$

7. Sum of squares within cells

$$\begin{aligned}
 SS_W &= SS_{\text{total}} - SS_{\text{teacher} \cdot \text{method}} \\
 &= 322 - 203.2 \\
 &= 118.8
 \end{aligned}$$

B. Summary of Analysis of Variance

Source of variation	df	SS	MS	F	Sig.
Teacher	(t-1)=1	2.133	2.133	0.431	n.s.
Method	(m-1)=2	180	90.00	18.182	.01
Interaction	(t-1)(m-1)=2	21.067	10.5335	2.128	n.s.
Within	tm(n-1)=24	118.8	4.95		
Total	=29	322.00			

Calculation of F values:

$$\begin{aligned}
 \text{For Teacher: } \frac{MS_{\text{Teacher}}}{MS_{\text{Within}}} &= \frac{2.133}{4.95} = 0.431 \\
 \text{For Method: } \frac{MS_{\text{Method}}}{MS_{\text{Within}}} &= \frac{90.00}{4.95} = 18.182 \\
 \text{For Interaction: } \frac{MS_{\text{Interaction}}}{MS_{\text{Within}}} &= \frac{10.5335}{4.95} = 2.128
 \end{aligned}$$

Interpretation: To determine the significance of these F values, we have to consult Table K with appropriate df. The following values are obtained from Table K:

for $df=(1, 24)$, value of F to be significant at .05=4.26
at .01=7.82

for $df=(2, 24)$, value of F to be significant at .05=3.40
at .01=5.61

In the last column of the table of Summary of ANOVA above, the interpretation of the F values has been given. The Teacher factor and the Interaction are not significant (n.s.) while only the Method factor is significant at .01 level. Hence, the three methods have been found to be significantly different in their effects on the criterion scores. This result is consistent with the one obtained in one-way ANOVA problem presented in a previous section in Table 11.1.

11.4.1 Effect of Introduction of a Second Factor

Since the data used in the one-way ANOVA problem was repeated here by introducing a second factor of 'teacher' and

<i>One-way ANOVA</i>			<i>Two-way ANOVA</i>		
<i>Source of variation</i>	<i>df</i>	<i>SS</i>	<i>Source of variation</i>	<i>df</i>	<i>SS</i>
Between sets (Methods)	2	180.00	Methods	2	180.00
Within	27	142.00	Teacher	1	2.133
Total	29	322.00	Interaction	2	21.067
			Within	24	118.80
			Total	29	322.00

thus increasing the number of cells from three to six, a comparison of the sums of squares obtained in the two situations will be revealing.

The following observations can be easily made:

1. The SS for Method factor (between sets) remains the same in both the situations.
2. The SS for Within in the One-way situation has been further broken down into three components—SS for teacher, SS for interaction and SS within. If we add up these three sums of squares from the two-way situation, we obtain a value equal to the SS within of the one-way situation.
3. Addition of a second or further factor leads to the reduction of the value of SS_{within} and consequently of error variance. Since the error variance is the denominator of the F-ratio, the values of F will go up and thus increase the likelihood of the rejection of the null hypothesis.
4. The SS_{total} remains the same in both the situations.
5. Reduction in error variance increases the precision of the experimental results.
6. The degrees of freedom for within variance in the one way problem have also been shared by the additional factor of teacher, interaction and within, in the two-way problem. The total df remains the same in both the situations.
7. However, selection of the additional factors for introduction in the experiment should be done carefully. Otherwise, a lot of labour and expense will go waste. Only those factors which have a known relationship with the criterion may be used for the purpose.

11.5 Notation for Three-way Anova

The notation and procedure of calculation used in the two-way analysis of variance problem can be extended for application to three-way or larger designs of ANOVA. However, the notation for a three-way analysis problem is presented below.

The factorial design may also be diagrammed as under:

	Method ₁		Method ₂		Method ₃	
	Teacher ₁	Teacher ₂	Teacher ₁	Teacher ₂	Teacher ₁	Teacher ₂
School ₁	n = 5					
School ₂						
School ₃						

The various sources of variation are:

Main effects

Source

*df**

Method (M)

$m - 1$

Teacher (T)

$t - 1$

School (S)

$s - 1$

Interactions

Method X Teacher

$(m - 1)(t - 1)$

Method X School

$(m - 1)(s - 1)$

Teacher X School

$(t - 1)(s - 1)$

Method X teacher X School

$(m - 1)(t - 1)(s - 1)$

Within cells

$mts(n - 1)$

Total

$N - 1$

*The small letters, m, t, s, stand for number of levels of factors of method, teacher and school, respectively; n means No. of cases in each cell.

Notation

The notation for the three way analysis of variance design is given below. The meaning of small letters *m*, *t*, *s*, and *n* have already been given above.

$$\text{Correction Factor, } C = \frac{(\Sigma X_{mn})^2}{N}$$

$$SS_{Method} = \frac{\Sigma (\Sigma X_m)^2}{nts} - C$$

$$SS_{Teacher} = \frac{\Sigma (\Sigma X_t)^2}{nms} - C$$

$$SS_{School} = \frac{\Sigma (\Sigma X_s)^2}{nmt} - C$$

$$SS_{M:T} = \frac{\Sigma (\Sigma X_{mt})^2}{ns} - \frac{\Sigma (\Sigma X_m)^2}{nts} - \frac{\Sigma (\Sigma X_t)^2}{nms} - C$$

$$SS_{M:s} = \frac{\Sigma (\Sigma X_{ms})^2}{nt} - \frac{\Sigma (\Sigma X_m)^2}{nts} - \frac{\Sigma (\Sigma X_s)^2}{nmt} - C$$

$$SS_{T:s} = \frac{\Sigma (\Sigma X_{ts})^2}{nm} - \frac{\Sigma (\Sigma X_t)^2}{nms} - \frac{\Sigma (\Sigma X_s)^2}{nmt} - C$$

$$SS^o_{MTS} = \frac{\Sigma (\Sigma X_{mts})^2}{n} - C$$

$$SS_{M,T,S} = SS_{MT} + SS_{M:s} + SS_{T:s} + SS_{M:T} + SS_{M:s} + SS_{T:s}$$

$$SS_{within} = \Sigma X_{mts}^2 - \frac{\Sigma (\Sigma X_{mts})^2}{n}$$

$$SS_{total} = \Sigma X_{mts}^2 - C$$

*SS between cov. of *M*, *T*, and *S*. Direct calculation of $SS_{M,T,S}$ interaction was not possible except by subtracting SS for all main effects and interactions from total SS.

The student should understand the sample rules on which the above notation has been based.

Σ —Sum of

X—Any raw score or observation

$X\Sigma_{mts}$ = Sum of all X scores taken over all combinations of factors M, T and S. In fact, the correct term should

have been $\sum_{m=1}^m \sum_{t=1}^t \sum_{s=1}^s X_{mts}$ —However,

to avoid complications, multiple signs of summation with their limits have been left out.

ΣX_m = Sum of all X scores within various levels of M. Hence, there will be as many sums as there are levels of M factor.

ΣX_{mt} = Sum of X scores within various cells composed of Method and Teacher factors. If $M=3$; $T=2$; there will be 3×2 such cells and hence six sums.

Denominators of the terms can be mechanically determined. This will be the product of small letter not used in the numerator. For example, in $SS_{Method} = \frac{\Sigma(\Sigma X_m)^2}{nts}$, the small letter m has been used in the numerator leaving n, t and s for insertion into the denominator. Further more, in $SS_{MXS} = \frac{\Sigma(\Sigma X_{ms})^2}{nt}$, m and s have been used in the numerator, leaving n and t for the denominator.

These small letters as explained already stand for:

m = number of levels of factor M or Method factor

t = number of levels of factor T or Teacher factor

s = number of levels of factor S or School factor

n = number of observations in each cell.

11.6 Interaction

In any discussion, of analysis of variance, a consideration of the meaning and interpretation of the interaction of variables or factors becomes important. These interactions may be between two or more independent variables and may cloud the main effects and make the interpretation of their significance difficult. The interaction of Method and Teacher on the criterion

variable of achievement may be encountered if one teacher is more efficient on one method while the other teacher is more efficient on the other method. To illustrate, let M_1 and M_2 represent two methods of teaching a foreign language. Let T_1 and T_2 represent the two teachers participating in the experiment. Hypothetical mean achievement scores of three experiments are given below:

TABLE 11.4

Interaction of Method and Teacher (Hypothetical Mean Achievement Scores)

(A)				(B)			
	M_1	M_2	Av.		M_1	M_2	Av.
T_1	16	4	10	T_1	16	6	11
T_2	4	12	8	T_2	6	4	5
Av.	10	8	9	Av.	11	5	8

(C)			
	M_1	M_2	Av.
T_1	16	8	12
T_2	12	4	8
Av.	14	6	10

In Table 11.4 (A), it is obvious that M_1 (with an average of 10) is better than M_2 (with an average of 8). But on closer inspection, one may see that M_1 gives better results than M_2 when used by T_1 .

In Table 11.4 (B), M_1 results are better than those of M_2 , but M_1 is not nearly as effective with Teacher₁ as it is with Teacher₂.

In Table 11.4 (C), M_1 is equally more effective than M_2 with Teacher₁ and Teacher₂.

These situations are diagrammed below in Figure 11.1

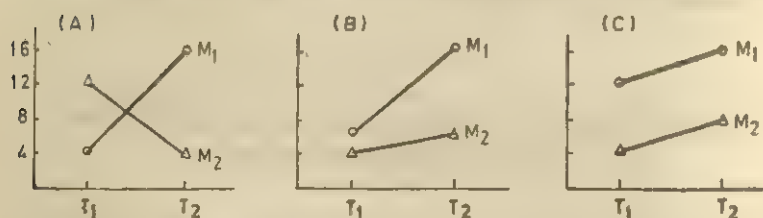


Fig. 11.1. Geometrical representation of the interactions based on the data of Table 11.4.

Figures 11.1 (A) and 11.1 (B) indicate presence of interaction in the two factors of method and teacher. In Figure 11.1(A), the two graphs intersect each other. It is called as *disordinal* interaction and is very difficult to interpret. The interaction represented by non-parallel and non-intersecting lines as in Figure 11.1 (B) is called *ordinal* interaction. Figure 11.1(C) shows parallel lines or lines with equal slopes, and represent absence of any interaction. Interaction occurs when the vertical differences between the lines in the geometrical representation are not equal. In the case of disordinal interaction, the direction of these differences is reversed from one level to the other. In the latter case, one cannot say that either method is superior to the other without qualifying the statement as to where the superiority lies in relation to the other variable. It is, therefore, advisable, for a clearer understanding, to draw the interactions geometrically.

Interactions can also be shown by subtracting the main effects of the independent variables from the resulting means. We give two illustrations of the procedure in Table 11.5.

In Table 11.5 (A.1), Teacher effects for T_1 and T_2 are $10 - 9 = 1$, and $8 - 9 = -1$ respectively. (Take the difference of the marginal averages from the general mean). If we subtract

TABLE 11.5

Subtraction of the Main Effects*A. Interaction Exists*

(A.1) Original Data

	M ₁	M ₂	Av.
T ₁	16	4	10
T ₂	4	12	8
Av.	10	8	9

(A.2) Teacher Effects
subtracted

	M ₁	M ₂	Av.
T ₁	15	3	9
T ₂	5	13	9
Av.	10	8	9

(A.3) Method and Teacher effect subtracted

	M ₁	M ₂	Av.
T ₁	14	4	9
T ₂	4	14	9
Av.	9	9	9

B. No Interaction Exists

(B.1)

	M ₁	M ₂	Av.
T ₁	16	8	12
T ₂	12	4	8
Av.	14	6	10

(B.2)

	M ₁	M ₂	Av.
T ₁	14	6	10
T ₂	14	6	10
Av.	14	6	10

	B.3		
	M ₁	M ₂	Av.
T ₁	10	10	10
T ₂	10	10	10
Av.	10	10	10

the first effect, 1, from all averages in the first row and add 1 to all averages in the second row, we have Table 11.5 (A.2). Similarly, we can take out the main effects due to method by subtracting $10 - 9 = 1$ from the first column and adding 1 to the second column. Table 11.5 (A.3) gives the resultant averages which show the direction of the interaction. Moreover, statistical tests can also be used to test their significance.

In Table 11.5B, the process of subtracting the main effects from the data of Table 11.4 (C) has been demonstrated with the result that in Table 11.5 (B 3) all averages have become equal to the grand mean. This shows absence of any interaction.

It may, however, be remembered that the graph or adjusted averages can give us an idea only of the presence or absence of an interaction. But, whether the interaction was significant or not would require an appropriate test of significance.

11.7 Assumptions Underlying the Analysis of Variance

As is the case of all parametric statistical tests, in the mathematical development of the analysis of variance, a number of assumptions have been made. It is important to take a look at the procedures of collecting data, and the nature of the distribution of the data obtained before taking a decision to use this technique. Normally, the data should satisfy the following assumptions:

1. *Normality of Distribution:* The dependent variable in the population from which the samples have been drawn should be normally distributed. Generally, a test of goodness of fit is used to ascertain the fulfilment of this pre-requisite. Violation of

this assumption will make the results appear somewhat more significant than they actually are. However, this deficiency can be made up by using a somewhat more rigorous level of confidence than usual. In case of very large N , near normality may be obtained without much difficulty.

2. *Homogeneity of Variance*: Homogeneity of variance means that the variances in the different sets of scores do not differ beyond chance. The $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ is

generally tested by using either Hartley's procedure or Bartlett's Test. Gross departures from homogeneity may lead to results which are seriously in error. However, mild departures from homogeneity of variance may not affect the results much. Transformation of scores to generate homogeneity or the use of a non-parametric test instead of ANOVA are generally recommended to avoid the violation of this assumption.

3. *Additivity of Effects*: As stated in an earlier section, the basic model of ANOVA states that a given observation or score is a sum of certain components each due to the effect of a particular identifiable source of variation. In most cases, this assumption is generally met.

4. *Random Sampling*: The sampling within the various sets should be random. It usually means that observations are mutually independent and with equal chance of selection.

11.8 General Uses and Limitations of Anova

Since the very inception of this wonder technique, the researchers have found it useful for the interpretation of experimental and observational data. It has been widely used in social and biological sciences. Since problems in these areas are generally multi-dimensional, ANOVA provides an appropriate and a powerful technique of analysis in which several factors can be simultaneously used and their effects tested. The research has now advanced from the single-variable classical experiments to the factorial designs in Fisher's tradition.

ANOVA, not only provides us an overall test of significance among several means, but also allows us to test for interactions which is not possible with t test. Sometimes only interactions are of interest to the researcher.

The rationale of ANOVA allows for the extension of the single classification model to a multiple classification model using three, four or more factors at a time. Thus, ANOVA has opened up the possibilities of research and statistical analysis in situations which require the use of multiple treatments

Several new experimental designs randomized blocks designs, repeated measures designs, Latin squares designs, Greco-Latin squares designs and a host of other designs have been devised.

Based on very sound mathematical assumptions and model, ANOVA provides a very powerful parametric test of significance. However, this technique suffers from the following *limitations*:

- (i) It is based on four rigorous assumptions. If these assumptions are not fully met by the data, the results may be in error.
- (ii) It may become difficult to interpret the results if triple or quadruple interactions turn out to be significant
- (iii) Set of data may not always be independent and thus involve correlated means.
- (iv) It imposes a very strict set of requirements in designing experiments, which may not be always possible to achieve.
- (v) The minimum number of cases in each cell should be 10 to have confidence in the results.
- (vi) ANOVA provides an overall test of significance among various means. We may need post ANOVA t test to locate the significant pairs of means.
- (vii) In case of comparison of two means, F test provides no additional information as compared with the t test. The relationship between the two in such situation is:

$$t = \sqrt{F} \quad (11.19)$$

$$\text{or } F = t^2 \quad (11.20)$$

Exercises for Practice

- 11.1 What do you mean by Analysis of Variance? Justify this nomenclature and give the theoretical rationale of ANOVA.

- 11.2 Compare F-ratio test and t test in terms of their relative merits and limitation. What is the mathematical relationship between the two?
- 11.3 State the assumptions of ANOVA. Discuss what happens when these assumptions are violated?
- 11.4 Apply ANOVA on the following sets of scores. State the H_0 and H_1 . Set the summary table and interpret the results.

(a) Set I	Set II	Set III	Set IV
3	5	8	2
4	5	7	3
6	5	8	1
8	5	7	
4		4	

(There are unequal n's in this case)

(b) Set I	Set II	Set III	Set IV	Set V
10	5	3	6	7
6	2	8	9	7
4	1	4	8	7
5	1	0	6	7
10	1	0	1	7

(Use deviation score method and also the raw score method. Do they give similar results?)

- (c) An experimenter wanted to see the relative effects of three drugs on the physical growth of rats. Three groups of rats were randomly selected from the same species. In each group, the number of male and female rats was kept equal. The gain in ounces in the weights of the rats is given below. By using an appropriate statistical technique, find out if the effects due to drugs, sex and the interaction of the two were significant:

	Drug 1	Drug 2	Drug 3
Male	2	5	4
	3	3	4
	1	3	4
	0	3	3
	2	4	5
Female	4	2	5
	3	2	5
	0	2	5
	0	1	3
	1	0	2

- (d) An experiment on the relative effectiveness of three different methods of teaching map-reading was conducted. Three teachers participated in the experiment. Groups of three students each selected randomly were assigned to various treatments. At the end of the experiment, criterion scores were obtained which are given below. Test the two main effects and interaction for significance.

	Method 1	Method 2	Method 3
Teacher 1	0	3	5
	1	5	1
	2	6	2
Teacher 2	1	0	0
	3	4	5
	5	0	7
Teacher 3	5	3	0
	3	3	4
	2	3	2

- (e) Four groups of 8 students each having an equal number of boys and girls were randomly selected and assigned to four different conditions of an experiment. Use ANOVA to test the main effects due to conditions and sex, and the interaction of the two:

	Condition 1	Condition 2	Condition 3	Condition 4
Boys	7	9	12	12
	0	4	6	14
	5	5	10	9
	8	6	6	5
Girls	3	4	3	6
	3	7	7	7
	2	5	4	6
	0	2	6	5

- 11.5 On the data of 11.4 (b) above, use post-ANOVA t test to locate the significant pairs of means.
- 11.6 How many degrees of freedom are associated with the variation in the data for:
- a comparison of four means for independent samples, each containing 15 cases.
 - a comparison of three means for independent samples, each containing 10 cases.
 - in a $3 \times 2 \times 4$ factorial experiment with three factors A, B and C, and $n=5$, all independent samples.
- 11.7 State five situations from education and psychology in which the use of ANOVA can be recommended.
- 11.8 What are the various limitations of ANOVA?
- 11.9 What is an 'Interaction'? Define ordinal and disordinal interactions. Draw the graphs of the following data:

	(A)			(B)	
	M_1	M_2		M_1	M_2
S_1	18	10	Boys	10	6
S_2	16	8	Girls	2	10

Subtract main effects from each table and show the presence or absence of interaction in the adjusted means.

CHAPTER 12

THE ANALYSIS OF COVARIANCE, ANCOVA

12.1 Introduction

Experimental designs, very often, require control of the intervening variables so that the results observed can be attributed, within certain limits of sampling error, to the treatment variable and to no other causal factor. Random assignment of subjects to various experimental treatments and matching of subjects for making up equivalent groups are two important procedures for this purpose. However, the experimenter may fail to control one or two conditions experimentally due to administrative difficulties or through ignorance of their relationship with the criterion measures. In such cases, where experimental control of a covarying variable or covariate has not been done, analysis of covariance (ANCOVA) provides a method of statistical control of the differential in the criterion scores attributable to the covariate.

For illustration, suppose an experiment is conducted to compare the relative effectiveness of traditional method, programmed learning method, and multi-media method in teaching geography. Random formation of three groups of students for instructional purposes was not possible. Hence, intact classes were used. These classes may differ in intelligence which is an important and known covariate or correlate of academic achievement (The Criterion). Intelligence, thus remained an uncontrolled variable. If intelligence scores of the groups were available from the records or could be obtained otherwise, analysis of covariance provides a method of statistical control of the variation or differential due to intelligence; the adjusted means could then be compared meaningfully.

Analysis of covariance which is an extension devised by R.A. Fisher, of his methods of analysis of variance, enables us even to dispense with the inconvenient procedure of matching of groups and secure the same increase in precision by the use of statistical controls. The hypothesis to be tested is that there are no differences in the various treatments, and that any differences in final criterion mean scores of the treatment groups, after allowances have been made for chance differences in the covariate mean scores, are due entirely to chance fluctuations in random sampling. The allowances for initial differences are to be made in terms of the regression of criterion measures on covariate measures. Under ANCOVA, it is assumed that there is homogeneity of regression which means that there is one true regression of criterion scores on covariate scores which is the same for all the groups.

To sum up, the method of analysis of covariance enables us:

1. to estimate the true regression of criterion scores on covariate scores with the assumption that there is no real difference in regression from group to group.
2. to use this regression coefficient to correct or 'adjust' the criterion means so as to allow for differences in the covariate measures, and
3. to test the differences remaining in the adjusted means.

In a one-way analysis of variance for comparing means, each criterion score, Y , can be expressed as

$$Y = \mu + \text{treatment effect} + \text{error} \quad (12.1)$$

If the purpose is to separate differences due to covariate from the criterion differences and do an analysis of covariance, the model changes. The covariate score (in our example intelligence) is denoted by X . It is related to the criterion measure, Y , by the regression equation,

$$Y - M_Y = \beta (X - M_X) \quad (12.2)$$

β here denotes, regression coefficient.

Model for ANCOVA, then would be

$$Y = \mu + \text{treatment effects} + \beta(X - M_X) + \text{error} \quad (12.3)$$

In one-way classification with two covariates, the model extends to

$$Y = \mu + \text{treatment effects} \\ + \beta_1(X_1 - M_{X_1}) + \beta_2(X_2 - M_{X_2}) + \text{error} \quad (12.4)$$

Increase in precision of results of ANCOVA depends on the degree of correlation between criterion (Y) and covariate (X). Some statisticians say that ANCOVA will result in no appreciable change in the adjusted means if this correlation is below .60.

12.2 Computation

In order to illustrate the computational procedure of analysis of covariance, let us assume that three methods of teaching English as a foreign language are applied to three randomly chosen groups of five subjects each and the criterion measures, Y, are obtained. From the school records, their intelligence scores or covariate measures, X, are also taken. The scores and the procedural steps in the analysis of covariance are given in Table 12.1 on page 273.

Step 1: Correction terms $C_X = (65)^2/15 = 281.67$

$$C_Y = (85)^2/15 = 481.67$$

$$C_{XY} = (65 \times 85)/15 = 368.30$$

Step 2: Total SS For $X = 369.00 - 281.67 = 87.33$

$$Y = 767.00 - 481.67 = 285.33$$

$$XY = 438.00 - 368.33 = 69.67$$

Step 3: Between Group Mean SS

$$\text{For } X = \frac{20^2 + 30^2 + 15^2}{5} - 281.67 = 23.33$$

$$Y = \frac{25^2 + 50^2 + 10^2}{5} - 481.67 = 163.33$$

$$XY = \frac{(20 \times 25) + (30 \times 50) + (15 \times 10)}{5} - 368.33 \\ = 61.67$$

TABLE 12.1

Worksheet for Covariance Analysis

Group I					Group II				Group III					
X_1	Y_1	X_1Y_1	X_1^2	Y_1^2	X_2	Y_2	X_2Y_2	X_2^2	Y_2^2	X_3	Y_3	X_3Y_3	X_3^2	Y_3^2
6	8	48	36	64	8	10	80	64	100	3	2	6	9	4
3	7	21	9	49	6	15	90	36	225	6	2	12	36	4
8	5	40	64	25	3	14	42	9	196	2	3	6	4	9
2	3	6	4	9	8	8	64	64	64	2	1	2	4	1
1	2	2	1	4	5	3	15	25	2	2	2	4	4	4
Sums														
20	25	117	114	151	30	50	291	198	594	15	10	30	57	22
M's														
4	5				6	10				3	2			

For all groups:

$$\Sigma X = 65; \Sigma Y = 85; \Sigma X^2 = 369; \Sigma Y^2 = 767; \Sigma XY = 438.$$

Step 4: Within Groups SS For $X=87.33-23.33=64.00$

$$Y=285.33-163.33=122.00$$

$$XY=69.67-61.70=8.00$$

Step 5: Analysis of Variance of X and Y scores taken separately.

TABLE 12.2

Summary of ANOVA

Source of variation	df	SS_X	SS_Y	MS_X	MS_Y
Between Means	2	23.33	163.33	11.67	81.67
Within Groups	12	64.00	122.00	5.33	10.17
Total	14	87.33	285.33		

$$F_X=11.67/5.33=2.18; F_Y=81.67/10.17=8.03.$$

From Table K, F values with df (2, 12); at .05=3.88.

$$.01=6.93.$$

F for Y is significant at .01 level. F for X is not significant which shows that there are no significant differences among the covariate, X, means. Hence the experimenter was successful in getting random samples in Groups I, II and III.

Step 6: Computation of Adjusted SS for Y, i.e. $SS_{Y.X}$

$$\text{Total } SS_{Y.X}=285.33-\frac{(69.67)^2}{87.33}=229.75$$

$$\text{Within } SS_{Y.X}=122.00-\frac{(8)^2}{64.00}=121.00$$

$$\text{Between } SS_{Y.X}=229.75-121=108.75$$

TABLE 12.3

Summary of ANCOVA

Source of variation	df	SS _X	SS _Y	SS _{XY}	SS _{Y·X}	MS _{Y·X} (V _{Y·X})
Between Means	2	23.33	163.33	61.67	108.75	54.38
Within Groups	11*	64.00	122.00	8.00	121.00	11.00
Total	13	87.00	285.33	69.67	229.75	

*1 df lost because of regression of Y on X.

$F_{Y·X} = 54.38/11.00 = 4.94$ (significant at .05 level but not at .01 level)

$$SD_{r·X} = \sqrt{V_{Y·X}} = \sqrt{11} = 3.32$$

Step 7: Correlation and Regression

$$r_{total} = \frac{69.67}{\sqrt{87 \times 285.33}} = .44$$

$$r_{between\ means} = \frac{61.67}{\sqrt{23.33 \times 163.33}} = .98$$

$$r_{within\ groups} = \frac{8.00}{\sqrt{64.00 \times 122.00}} = .09$$

$$b_{within\ groups} = \frac{8.00}{64.00} = .125$$

Step 8: Calculation of Adjusted Y Means

TABLE 12.4

Adjusted Y Means

Group	N	M_X	M_Y	$M_{Y·X}$ (Adjusted)
I	5	4	5	5.04
II	5	6	10	9.79
III	5	3	2	2.17
General Mean		4.33	5.67	

$$M_{Y.X} = M_Y - b(M_X - GM_X) = M_Y - bx \quad (12.5)$$

(Y means adjusted for X variations)

b: denotes regression coefficient,

x: denotes deviation of M_X from GM_X

other symbols are as before.

For Group I: $M_Y - bx = 5 - .125(4 - 4.33) = 5.04$

For Group II: $M_Y - bx = 10 - .125(6 - 4.33) = 9.79$

For Group III: $M_Y - bx = 2 - .125(3 - 4.33) = 2.17$

Step 9: Comparison of Adjusted Means

Post ANCOVA t test can be used for the purpose

Adjusted MS_W for Y.X is the estimate of the population variance available. Hence

$$SD_{Y.X} = \sqrt{11} = 3.23$$

$$SE_{M_D} = \sigma_W \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (12.6)$$

$$= 3.23 \sqrt{\frac{1}{5} + \frac{1}{5}}$$

$$= 2.04$$

$$t \text{ values: } M_1 - M_2 = \frac{5.04 - 9.79}{2.04} = -2.33 \text{ Significant at } .05.$$

$$M_1 - M_3 = \frac{5.04 - 2.17}{2.04} = 1.41 \text{ not significant}$$

$$M_2 - M_3 = \frac{9.04 - 2.17}{2.04} = 3.73 \text{ Significant at } .01.$$

12.3 Notation and Description of Computational Steps

The actual computational steps with results have been presented above. However, for a better understanding, the same are described below and the relevant notation and formulas presented to outline the procedure.

Step 1: Correction term (C)

There are now three sets of data in each group—Covariate

X; Criterion Y; and Cross products XY. Keeping this fact in view, we need three correction terms:

$$\text{Correction for X, } C_X = \frac{(\sum X)^2}{N} \quad (12.7)$$

$$\text{Correction for Y, } C_Y = \frac{(\sum Y)^2}{N} \quad (12.8)$$

$$\text{Correction for products, } C_{XY} = \frac{\sum X \times \sum Y}{N} \quad (12.9)$$

Step 2 SS for totals

We have three SS's for total: Sums of squared scores over all the groups are used:

$$\text{For X} = \sum X^2 - C_X \quad (12.10)$$

$$\text{For Y} = \sum Y^2 - C_Y \quad (12.11)$$

$$\text{For XY} = \sum XY - C_{XY} \quad (12.12)$$

SS_{XY} were obtained by multiplying pairs of X and Y scores, summing over the entire range and subtracting C_{XY} : thus $(6+8+3 \times 7 + \dots + 2 \times 2)$. In the worksheet col. 3 of each group shows the cross products which were summed up over all the three groups

Step 3 SS between means

SS between means for X and for Y follow the method of ANOVA while SS between means for XY is the sum of the corresponding X and Y column totals divided by n i.e., 5, and minus C_{XY} .

$$\text{Thus SS between means for XY} = \frac{\sum (\sum X \times \sum Y)}{n} \quad (12.13)$$

in which k stands for number of groups, which is three in this case.

Step 4 SS within groups

These SS's for X, Y and XY were found by subtracting between mean SS's from total SS's.

For X = Total SS for X — between means SS for X

For Y = Total SS for Y — between means SS for Y

For XY = Total SS for XY — between means SS for XY .

Step 5 Preliminary Analysis of Variance

A preliminary analysis of variance on X and Y scores separately was done and results presented in a summary table. The groups did not differ significantly on the covariate, X . However, there were significant differences among the Y means. This analysis is done to have a pre-adjustment view of the differences in the criterion means and also to see the differences in the covariate over the various groups.

Step 6 Adjusted SS for Y

This step is taken to remove from Y scores, any variability contributed by X scores. The adjusted $SS_{Y.X}$ are symbolized as $SS_{Y.X}$. The generalized formula is

$$SS_{Y.X} = SS_Y - \frac{(SS_{XY})^2}{SS_X} \quad (12.14)$$

(Sum of squares of Y adjusted for X differences)

Thus, $SS_{Y.X}$ total would be given by

$$\text{Total } SS_{Y.X} = \text{Total } SS_Y - \frac{\text{Total } SS_{XY}^2}{\text{Total } SS_X} \quad (12.15)$$

and

$$\text{Within } SS_{Y.X} = \text{Within } SS_Y - \frac{\text{Within } SS_{XY}^2}{\text{Within } SS_X} \quad (12.16)$$

Between means $SS_{Y.X}$ cannot be readily calculated directly. Hence these are to be obtained by subtracting Between means $SS_{Y.X}$ from Total $SS_{Y.X}$. Thus,

$$\text{Between Mean } SS_{Y.X} = \text{Total } SS_{Y.X} - \text{Within } SS_{Y.X} \quad (12.17)$$

The variances ($MS_{Y.X}$) were computed from the various adjusted sums of squares by dividing the latter by appropriate

df. Owing to the adjustment of Y scores, and the additional restriction imposed by formula (12.14), 1 df was lost, thus giving a total df equal to 13 instead of 14; and df for 'Within groups' equal to 11 instead of 12.

A comparison of the F_Y (pre-adjustment) and $F_{Y.X}$ (Adjusted) would show that the former value of F which was 8.03 and significant at .01 level has come down to 4.94 which is not significant at .01 level but only at .05 level. Hence, one can see the utility of using ANCOVA. The pre-adjustment Y mean differences have become less sharp after adjustment of the variability due to the covariate, X. This is evident from an inspection of the values of adjusted means. The smallness of changes is due to the fact that the correlation between criterion and covariate scores for total is .44 only and r for within groups is very small i.e., .09 (See step 8).

Step 7 Correlation and Regression

This step is taken to obtain various correlations and regression coefficients for total, between means and within groups. This helps in a better understanding of the results of step 6 above. The general formula used is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} \quad (12.18)$$

The regression coefficient, b, is given by

$$b = \frac{\sum xy}{\sum x^2} \quad (12.19)$$

$\sum xy$, $\sum x^2$, $\sum y^2$: denote sums of squares, respectively for cross products, x and y.

This may be applied to the appropriate SS's for total, between means and within groups.

The Within-groups correlation is a better measure of the relationship between the intelligence scores, X, and the achievement scores, Y, than the total correlation as the systematic differences have been removed from the within r. This correlation is very small, rather negligible and hence did not lead to much reduction between the Y means when variability due to X

is kept constant. High correlation "between means" has reduced the numerator of the $F_{X,Y}$ ratio, from $SS_Y=163.33$ to $SS_{Y.X}=108.75$ while the very low correlation within groups has almost not reduced the denominator of the variance ratio, $F_{X,Y}$ as within groups sum of squares has come down only by one unit, i.e., from $SS_Y=122$ to $SS_{Y.X}=121$. When correlation among scores is high and correlation among means is low, non-significance of results of preliminary ANOVA is likely to change into significance on adjustment of SS. However, if the case is reverse as in our example, the value of F in ANOVA is likely to come down in ANCOVA.

Regression coefficient, b_{within} was calculated as a nearly unbiased estimate of the regression of Y on X , since any systematic influence due to differences among means has been removed. Therefore, this coefficient was used in the calculation of adjusted Y means.

Step 8 Adjusted Means

After ANCOVA has been completed, criterion mean scores (Y means) can be adjusted directly for differences in the X means. The formula for the same is $M_{Y.X}=M_Y-b(M_X-GM_X)$, using regression coefficient b_{within} which in our example is .125. For each group, the same formula is to be used. M_X stands for the mean of X scores of the corresponding group whose Y mean is being adjusted. GM_X is the general mean of X scores which remains the same for all groups. The adjusted mean as obtained in our example have not changed much because of the reasons explained in step 7 above.

Step 9 Comparison of Adjusted Y means

As a post-ANCOVA step, significant pairs of adjusted Y means can be determined by using formula (11.12) and the procedure discussed in Chapter 11. For this purpose, within $MS_{Y.X}$ from the ANCOVA table is used as an estimate of the population variance and $SD_{Y.X}$ can be readily calculated by taking its square-root. The procedure has been explained in the computational step No. 9. The estimate of the population variance is readily available in the form of MS_{within} for adjusted

Y scores from the ANCOVA table. Two pairs of means — $-M_1 - M_2$; and $M_2 - M_3$ turn out to be significant.

12.4 Assumptions Underlying Ancova

Analysis of covariance is based on some assumptions, warranted by the mathematical logic on which it is based. These assumptions include the usual four assumptions made for analysis of variance and two additional ones. These are:

1. Normality of distribution,
2. Homogeneity of variance,
3. Additivity of effects,
4. Mutual exclusiveness and independence of observations which may in other words be termed as random selection.
5. A linear relationship between X and Y.
6. Homogeneity of regression or similarity of the slope of the regression lines for each experimental group. For this purpose $H_0: \beta_1 = \beta_2 = \dots = \beta_k$ is tested by using an F test. This F-ratio is based on the difference between sum of squares for pooled and separate regression lines and sum of squares for separate regression lines.

12.5 General Uses of Ancova

The basic model of analysis of covariance can be further extended to include a double classification in which an experiment is duplicated in randomly selected situations. For example, the experiment described in section 11.2 can be repeated in a number of randomly selected schools. The procedure in this case will be much the same as before. The principal difference, however, will be that the error estimate will be based on the interaction (Method \times School) variance, rather than on the "Within groups" variance of the adjusted scores. The 'adjustment' of the criterion scores will accordingly be based on a regression coefficient derived from the sum of squares of Method \times School interaction rather than for "Within groups".

If it is desired to eliminate the effect of more than one uncontrolled variables in the experiments of the type described

in this chapter, it may be done by an extension of the methods described in the preceding pages. The use of multiple regression equation will be required for adjustment of criterion means by eliminating the influence of the uncontrolled variables. As the number of covariates increases, the computational labour will be tremendous. However, the advantage gained depends upon the magnitude of the multiple correlation coefficient. Hence, it is advisable to select the best combination of variables. Some of the uses of ANCOVA were pointed out in the section on introduction. Some other uses of ANCOVA are given below:

1. It leads to increase in the precision of randomized designs or experiments. It is very useful in situations where a chance assignment of a particular treatment to group already better than other groups has taken place.
2. It allows for the adjustment of sources of bias in observational studies. For example, the differential effects of covariates of age and weight can be eliminated from the dependent variable of time taken in covering a distance.
3. It allows for probing into the nature of treatment effects after significant main effects have been obtained.
4. It permits, as a by-product, the study of regression in multiple classification.

To conclude, analysis of covariance is a powerful technique for removing bias, effects of disturbing variables and environmental sources of variations. However, this holds good when the assumptions of ANCOVA are met and a good combination of variables achieved, and the interaction between the treatment and covariate is absent. ANCOVA, in no way, be taken as an answer or a substitute for ability to randomize and provide an experimental control. Stratified random sampling is preferable to ensure pre-experimental equivalence of groups and thus to eliminate the need of using ANCOVA.

Exercises for Practice

12.1 Three intact groups were taken and randomly assigned to three treatments. Pretest and post-test measures were obtained as follows:

<i>Treatment A</i>		<i>Treatment B</i>		<i>Treatment C</i>	
<i>Pre-test</i>	<i>Post-test</i>	<i>Pre-test</i>	<i>Post-test</i>	<i>Pre-test</i>	<i>Post-test</i>
<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>
10	15	16	15	15	22
12	19	18	18	15	21
16	20	17	16	14	20
11	18	15	14	13	18
11	18	14	17	13	19
Sums 60	90	80	80	70	100
M's 12	18	16	16	14	20

- (a) Do analysis of variance and then analysis of covariance. Compare the results.
 - (b) Explain the possible reasons for the differences in results.
- 12.2 On what single factor does the gain in precision in ANCOVA depend.
- 12.3 If the covariate mean is greater than the X grand mean in one group, what would be the effect on the adjusted Y means for that group?
- 12.4 Why would a randomized-block-analysis be more robust than an ANCOVA?
- 12.5 State and explain the assumptions underlying ANCOVA.
- 12.6 Give some uses of ANCOVA in educational and psychological research.

CHAPTER 13

RELIABILITY AND VALIDITY OF TEST SCORES

13.1 Reliability

The term "reliability" as used in everyday language conveys a meaning which is somewhat parallel to the meaning ascribed to it by the measurement expert. Suppose we say that a worker was reliable. What does this convey? It might be supposed that the worker reports at the same time everyday, appears in the same condition, and performs consistently. This similarity of behaviour from time to time would be comparable to one of the tester's approaches to reliability. "Trustworthiness" is another word used by the common man to signify the term reliability of human beings. Reliability pertains to a class of test characteristics. These are stability, equivalence and internal consistency. These are time-associated and form-associated bases of reliability. Reliability is not concerned with the appropriateness of measurements, an issue that falls within the scope of the meanings of another test characteristic called validity. The main concern of reliability is an accurate repeatability of scores over time and parallel forms of a test. Reliability of physical measurements using steel tape, etc. is very high as these instruments give consistent results with high accuracy.¹ However, psychological tests and instruments are less reliable as many measurement errors are bound to creep in. The *coefficient of reliability* of a test forms the basic index of reliability reported in the literature. It is based on self correlation of a test.

13.1.1 Methods of Estimating Reliability

There are basically four different procedures generally used for computing reliability:

1. Test-retest method (Repetition of the same test or measure over time)
2. Alternate or parallel forms method (Administration of a second 'equivalent' form of the test)
3. Split-half technique (Sub-division of the test into two or more equivalent fractions)
4. Rational equivalence method.

All these methods furnish estimates of the reproducibility of test scores and one method may be preferred to another according to the demands of the structure. These methods are described below:

13.1.1.1 Test-Retest Method

This method involves (i) repetition of a test on the same group immediately or after a lapse of time, and (ii) computation of correlations between the first and second set of scores. The correlation coefficient thus obtained indicates the extent or magnitude of the agreement between the two sets of scores and is often called the *coefficient of stability*. Although test retest is sometimes the only available method, this procedure is open to several serious objections. Immediate repetition of a test may involve (i) immediate memory effects, (ii) practice effects, and (iii) confidence effects induced by familiarity of contents. Intervals of six months or longer in young children may show "growth effects". The factors of intervening learning and unlearning may lead to lowering of self correlations. It may not be possible to control conditions on the second testing.

Memory, practice and other carry over effects may be offset by increasing the time interval between the two testings.

13.1.1.2 Alternate or Parallel form method

This method involves the administration of two equivalent or parallel forms of the test instead of repetition of a single test. This avoids, to a great extent, the disadvantages of the test-retest method involving short or long intervals of time. The two equivalent forms are so constructed as to make them similar (but not identical) in content, mental processes involved, number of items and the difficulty levels of the items. The

subjects take one form of the test and then, as soon as possible, the other form. The correlation coefficient between the two sets of scores determines the agreement and is generally called as *coefficient of equivalence*. Alternate forms should be drawn very carefully by matching test material for content, difficulty and form. An interval of atleast two to four weeks should be allowed between the two administrations of the test, to offset the carryover effects due to familiarity of content. The method has some other limitations also. The second form of a test may not be available. A second testing may place heavier demands on the time and resources of the researcher and the subjects. The psychometricians have devised another technique called the split-half method which takes care of some of these defects.

13.1.1.3 The Split-half method

The most widely used procedure for estimating reliability from a single testing divides a particular test into two presumably equivalent halves. The test is divided into two halves only for the purpose of scoring and not for administration. It means that a single test is given at a single sitting and with a single time limit. However, two separate scores are derived - one by scoring one half and the other, by scoring the other half. The correlation between these two sets of scores provides a measure of the accuracy with which the test is measuring the individual. A sensible procedure generally used for splitting the test into two halves is the odd-even split technique. The odd numbered items, 1, 3, 5, 7, etc.; and the even numbered items, 2, 4, 6, 8, etc. form two different sets of items for scoring. This procedure is better than others like first half and the second half and split into blocks of 5 or 10 items etc. and ensures balancing out of factors of items form, content covered, and difficulty level in tests having 60 or more items.

The computed correlation, in this technique, is between two half-length tests. This value is not directly applicable to the full length test which is the actual instrument prepared for use. Hence Spearman Brown Prophecy Formula is used to estimate the reliability of the full length test from the self correlation of the half-tests.

$$r_{11} = \frac{2r_{1/2, 1/2}}{1 + r_{1/2, 1/2}} \quad (1)$$

(Spearman Brown Prophecy Formula for estimating reliability from two comparable halves of a test.)

where, r_{11} = reliability coefficients for the full length test

$r_{1/2, 1/2}$ = reliability coefficient of the half test found experimentally.

For example, if correlation between the two halves of a test is .60, the reliability of the full test will be

$$r_{11} = \frac{2(0.60)}{1 + 0.60} = \frac{1.20}{1.60} = .75$$

This split-half method is regarded to be the best by many psychometricians. The main argument in favour of this method is the single shot approach, leading to no place for errors due to repetition and lapse of time. Convenience and saving of time and expense are some other important considerations.

However, this method suffers from some *limitations*.

- (i) It is not useful for speeded tests.
- (ii) Splitting of the test can be done in several ways, thus leading to the non-comparability of reliability indices from these.
- (iii) It is not so useful for tests having a small number of items i.e. less than twenty.
- (iv) Chance errors may affect scores on two halves equally and thus lead to higher correlations.

13.1.1.4 "Rational Equivalence" method

The three methods described above suffer from some limitations. The method of rational equivalence is an approach to avoid some of these objections. This method is based upon intercorrelations of the items in the test and the correlations of the items with the test as a whole. Several formulas for calculating reliability by this method have been suggested. However, the formula, given below, is the most popular:

$$r_{11} = \frac{n}{(n-1)} \times \frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2} \quad (2)$$

(Kuder-Richardson Formula (20) based on difficulty and the intercorrelations of test items.)

in which,

r_{11} = reliability coefficient of the whole test

n = number of items in the test

σ_t = SD of the test scores

p = the proportion of the group answering a test item correctly

$q = (1 - p)$ = the proportion of the group answering a test item incorrectly.

Another formula which is a simple approximation to the above formula is as follows:

$$r_{11} = \frac{n \sigma_t^2 - M(n - M)}{\sigma_t^2 (n - 1)} \quad (3)$$

(Approximation to Formula 2)

where, r_{11} = reliability of the whole test

n = number of items in the test

σ_t = SD of the test scores

M = Mean of the test scores.

Formula (3) saves labour as it is based on the mean, SD and number of items. Correlation coefficient between two halves is not required.

These formulas are based on the assumption of equal difficulty of items which may not be generally met in power tests. The formulas provide a minimum estimate of reliability thus giving an underestimation.

13.1.2 Factors Affecting Reliability

13.1.2.1 Length of test

Other things being equal, the reliability of a test is a function of its length. Longer tests tend to be more reliable than shorter tests. Logically, the more samples we take of a given area of knowledge, skill, behaviour and the like, the more reliable will be our appraisal of that area. Lengthening the test means adding more items, or having several applications of the test or use of parallel forms. The Spearman Brown Prophecy Formula for calculating the reliability of a test with increased length is

$$r_{nn} = \frac{nr_{11}}{1 + (n-1)r_{11}} \quad (4)$$

in which, r_{nn} = the correlation between n forms of a test and n alternate forms

r_{11} = the reliability coefficient of Test 1.

Example: Suppose for a test of 50 items, the reliability coefficient is .80. What would be the reliability of a test if its length is doubled to 100 items; tripled to 150 items, and quadrupled to 200 items.

The calculations are as below:

first situation (Doubling the length)

$$r_{22} = \frac{2 \times 0.80}{1 + (2-1) \times 0.80} = .889$$

second situation (Tripling the length)

$$r_{33} = \frac{3 \times 0.80}{1 + (3-1) \times 0.80} = .923$$

third situation (Quadruplicating the length)

$$r_{44} = \frac{4 \times 0.80}{1 + (4-1) \times 0.80} = .941$$

Spearman-Brown formula shows that after we have reached a high degree of reliability, additional items do not improve the reliability enough to justify the extra time and effort required for building items for testing pupils. The increase in the values calculated above are not appreciable above the doubling of the test.

13.1.2.2 Range of Talent

The range of talent, achievement or ability of the pupils on whom the reliability is based has a direct effect on the reliability coefficients. The greater the variability in the group, the higher the reliability coefficient.

13.1.2.3 Testing Conditions

The conditions in which the test is administered and scored may affect reliability in either side. Other such factors are: mental set of pupils, level of their motivation, speed of work, emotional stability, distractions and accidents, and cheating by pupils

13.2 Validity

Validity of a test or evaluation device can be defined as the degree to which the test measures what it is intended to measure. A test which is meant to measure achievement in Mathematics should not measure achievement in History or any other subject. In physical sciences, validity of measuring instruments like scales, thermometers, chronoscopes and ammeters is found by comparing their measurements with standard measures. The validity of a psychological test is also found by comparison with standard (and sometimes arbitrary) measure. However, the validity of a physical instrument can be estimated accurately while it can never be done in the case of a mental test.

Validity is a relative term and has reference to a particular purpose or situation. The question "Is the test valid?" can be answered only by replying to the question "Valid, for what?" Hence, there are different types of validity meant for different purposes.

13.2.1 Types of Validity

13.2.1.1 Content validity

Content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn. It is based on a comparison of the analysis of test content with the analysis of the course content and the instructional objectives. It is seen as to how well the former represents the latter. The analysis is done essentially through logical, rational and judgemental process. That is why, sometimes the content validity may be referred to as *rational or logical validity*. Content validity is important primarily for measures of achievement. The test maker first determines the widely accepted goals of instruction in the subject and then prepares a blueprint for the test. Test content is drawn from the course content and weighted according to the weightage of the objectives of the course and the course content. An appraisal of the content validity of a test involves a careful and detailed examination of the actual test tasks.

13.2.1.2 Face Validity

Face validity has something to do with the mere appearance of a test. A test is said to have face validity when by appearance it "looks like" measuring what it is meant to measure. The appearance of reasonableness is spoken of as "face validity". The judgemental process is used to determine answers to such questions whether test content "appears to" correspond to that of the course. A test of Mathematics should have numerical questions; and a test of History, questions about kings, movements, wars, etc. The relevance of the test items to specific situations, age groups, language groups, etc. is also a matter of concern. However, no single numerical index of the face validity of a test can be calculated.

13.2.1.3 Concurrent Validity

Concurrent validity is evaluated by showing how well the test scores correspond to already accepted measure of performance or status made at *the same time*. For example, scores

on a test of knowledge of basic concepts in Geography can be validated against the teacher's ratings of the students on this aspect. Intelligence tests were first validated against school grades, teacher's ratings, etc. A newly constructed test of intelligence may be validated by finding its correlation with another already existing well accepted test in this area. In these cases, a correlation coefficient between the two sets of measures is calculated as an index of validity. The main problem is to set up a criterion which is independent and reliable.

13.2.1.4 Criterion related validity

In a discussion of concurrent validity as given above, the test is validated against a criterion at the same point of time. However, we may be interested in using a test to predict some future outcome. A test of aptitude for teaching may be used to admit students to teachers' training college and be expected to predict success at the job as teachers. A scholastic aptitude test may be used to predict how likely will the high school students be successful at a college. A clerical aptitude test may be used to predict success on the job as clerks. We are thus interested in success or performance in the future. This process is also called *Predictive Validity*. A correlation coefficient between the test scores and the criterion scores is calculated. The higher the correlation, the better the test as a predictor. However, the problem of selecting "an appropriate criterion" is very ticklish. The main problem arises when the criteria of success on the job is to be determined; the records are not available, time interval between completion of training and placement and working on the job for a period long enough to allow proper evaluation of success and the like. The success or failure of a worker may depend on conditions external to his own personality and skill. Ratings of success by superiors may be influenced by many factors other than the proficiency of the worker being rated. However, the following qualities are desired in a criterion measure:

- (i) Relevance
- (ii) Freedom from bias (providing equal opportunity to all to perform well.)

- (iii) Reliability, and
- (iv) Availability.

13.2.1.5 Construct validity

Sometimes questions like the following are asked, "What does this test mean or signify?" "What does the score tell us about the individual?" "Does it correspond to some meaningful trait or construct that will help us in understanding him?" These questions are related with the construct validity of the test. The term "construct" is used in psychology to refer to something that is not observable, but is literally "constructed" by the investigator to summarize or account for the regularities or relationships that he observes in behaviour. Thus, most names of traits refer to constructs. Intelligence, sociability, extraversion, aggressiveness, need-achievement and verbal reasoning are some examples of constructs. Tests of these functions are valid in so far as they behave in the way that such a trait should reasonably be expected to behave. A "theory" about a trait will lead to predictions of three types.

- (i) A theory may make predictions about correlations with other accepted measures of the function in question.
- (ii) It may make predictions about differences in the groups which are supposed to or known to possess the trait in high degrees and those in low or non-existent degrees just as delinquent and non-delinquents, intellectually superior and intellectually inferior groups.
- (iii) A theory may predict modification of a human characteristic as a result of certain experimental conditions or treatments.

For any test that presumes to measure a trait or quality, we can formulate a network of theory leading to definite predictions as explained above. Insofar as they are borne out, the validity of the test as a measure of the trait or construct is supported. In so far as the predictions fail to be verified, we are led to doubt the validity of our test or our theorizing, or both. Evidence of construct validity is partly rational and partly empirical and judgement and evidence join together in the validation enterprise.

13.2.1.6 Factorial validity

Factorial validity is, in a way, extension of the construct validity discussed above. The intercorrelations of a large number of tests are examined and if possible accounted for in terms of a much smaller number of more general "factors" or trait categories. Sometimes 3 or 4 factors may account for the intercorrelations among 15 to 20 tests. The factorial validity of a test is defined by its correlation with a factor, called factor loadings. A word comprehension test may correlate .82 with the verbal factor extracted from a test battery. This coefficient will then be the test's factorial validity.

13.2.2 Factors Affecting Validity

The following are some of the factors which affect test validity. The test users should recognize factors that tend to make their test invalid.

- (i) *Cultural factors* such as socio-economic status, social class structure, differential sex roles affect performance on various tests.
- (ii) *Response sets* or test-taking habits of the examinees may differentially affect the validity estimates.
- (iii) *An Increase in the Number of Test Items* may boost up reliability but may bring down the validity.
- (iv) *Difficult and Ambiguous Directions* to the pupils may render the test a measure of something different than the test author intended.

13.3 Relation Between Reliability and Validity

Reliability and validity are the two important aspects of the same quality of a test, called "test efficiency". Reliability is concerned with the stability of the test score and does not go beyond the test. Validity, on the other hand, implies evaluations in terms of outside—and independent-criterion. A test, to be reliable, need not be practically valid while a test to be valid must be reliable. A clock which gains twenty minutes a day is a perfectly reliable instrument as it will repeat the same gain every day. However, judged against a standard time piece, the clock is not valid.

13.4 Item Analysis

Item analysis of a test comes after the preliminary draft of a test has been constructed, administered on a group of students and scored. A tabulation is done to determine the following two important characteristics of each item.

- (i) Level of difficulty or item difficulty, and
- (ii) Discriminating power of the test items or item discrimination.

The above two indices help in item selection for the final draft of the test. Another step which precedes the calculation of item difficulty and item discrimination of a test is *item selection* based upon the judgement of competent persons as to the suitability of the item for the purposes of the test. This is, in brief, a step towards establishing the 'content validity' and face validity of the test items (already described). The procedure involves a referral of the items to the experts and obtaining of their consensus on each. There are several methods of item analysis described in various texts exclusively based on construction of tests. However, for the purpose of the present book, only a few generally more popular techniques will be presented.

13.4.1 Item Difficulty

Item difficulty can be gauged in various ways:

- (i) Expert ranking of the items in order of difficulty,
- (ii) Quickness by which the item can be solved, and
- (iii) Calculation of the proportion of students solving the item correctly.

While the first two are based on judgemental process, the last one is based on empirical evidence and generates a numerical index and is hence widely used.

Item difficulty may be defined as the proportion of the examinees that marked the item correctly. The numerical term which indicate the level of difficulty is called *difficulty index*. It may range between 0 and 100. An item answered correctly by 65%

students has a difficulty index of .65. If 90% of a standard group pass an item, it is easy; if only 10% pass the item is hard or too difficult. Generally, items of moderate difficulty (40–50–60% passing) are to be preferred to those which are much easier or much harder. If p is the proportion passing an item; q is $(1-p)$, or proportion failing, the SD of the item (its variability) is \sqrt{pq} and its variance (σ^2) is pq . The variance of an item is at its maximum when $p=q=.5$. Hence to bring out more individual differences (a greater spread), the item difficulty may be kept near .5.

Correction for Guessing

In multiple choice objective items guessing plays an important role and boosts up the result of those who do not understand the item nor know its correct answer and hence indulge in guesswork. Thus chance success must be corrected. A formula for the purpose is:

$$P_c = \frac{R - \frac{W}{(k-1)}}{N - HR}$$

in which,

P_c = the per cent of examinees who correctly know the answer (corrected index of item difficulty)

R = the number of examinees who get the right answer

W = the number of examinees who get the wrong answer

N = the number of examinees in the sample

HR = the number of examinees who do not reach the item and hence could not try solving it.

K = Number of alternatives in the item.

As an example, consider the following data: A sample of 200 students took a test of 100 items. Each item had 5 choices. Item number 28 was answered correctly by 60 and incorrectly by 96; 44 could not reach the item. Now the item difficulty can be calculated as below:

$$P_c = \frac{60 - \frac{96}{(5-1)}}{200 - 44} = \frac{60 - 24}{156} = 36/156 = .23$$

The uncorrected index of item difficulty would have been R/N or $60/200$ or $.30$. The use of the formula for correction for guessing has brought it down to $.23$.

The student should be aware of the fact that as the number of options or choices increases the effect of guessing on success decreases. This effect is the largest in a true-false item. The formula given above is based on two assumptions: (i) Wrong answers are due to lack of knowledge, and (ii) to an examinee who does not know the correct answer, all the options or choices are equally attractive.

13.4.2 Item Discrimination

Item discrimination or the discriminating power of a test item refers to the degree to which success or failure on an item indicates possession of the ability being measured. It determines the extent to which the given item discriminates among examinees in the function or ability measured by the item. The procedure involves the following steps:

- (i) Administration of the draft test on a sample of about 300.
- (ii) Identification of upper 27% and bottom 27% examinees. (having highest and lowest scores in rank order respectively on the total test).
- (iii) calculation, in respect of each item, of the percentage/proportion of the examinees attempting it correctly.
- (iv) The discrimination index, DI will be given by the following formula:

$$DI = P_u - P_L$$

in which, DI = discrimination index,

P_u = Proportion in the upper group passing the item,

P_L = Proportion in the lower group passing the item.

- (v) The DI can be tested for significance by using a critical ratio test and items with positive and significant differences retained.
- (vi) The value of the discrimination index can range from -1.00 through zero to $+1.00$.

- (vii) An item is said to have negative discrimination power if poor students answer it correctly more often than the good students or $P_u < P_L$.
- (viii) Items having negative discrimination are rejected. Items having discrimination index above .20 are ordinarily regarded satisfactory for use in most tests of academic achievement.

Several other indices of discrimination based on upper and lower 27% groups can be calculated. These include calculation of normalized biserial coefficients which can be readily read from a table prepared by J.C. Flangan.

The size of an acceptable item validity index will depend upon the length of the test, the range of difficulty indices, and the purposes for which the test has been designed. The poor items are removed or improved for inclusion in the final test.

Questions for Practice

- 13.1. What do you mean by Reliability? How is it established?
- 13.2. Must all tests be reliable?
- 13.3. What is the effect of lengthening of a test on reliability?
- 13.4. A test of 60 items has a reliability coefficient of .60. What will be the reliability coefficient of the test if
 - (i) the number of items is increased to 120
 - (ii) the number of items is increased to 180.
- 13.5. Estimate the reliability of a test of 100 items if $\sigma_e = 10.00$; $\Sigma pq = 20.00$. Which method of reliability is applicable here?
- 13.6. What does the term 'validity' stand for? Explain various types of validity that a test can have.
- 13.7. Given the following data for 4 items. Calculate difficulty and discrimination indices for each.

<i>Item No.</i>	<i>% right in upper 27%</i>	<i>% right in lower 27%</i>	<i>Difficulty Index</i>	<i>Discrimination Index</i>
10	80	60		
15	45	40		
21	50	26		
38	79	21		

CHAPTER 14

REGRESSION AND PREDICTION

Scientific investigations generally aim at one or more of the three important things—explanation, prediction and control. All the three goals of the scientist have their own importance in studying the natural, physical and social phenomena. However, one of the most important tests of any scientific hypothesis is its ability to make predictions. Statistical reasoning has helped the scientist in framing statements of a predictive nature. The amount of error in the predictions can also be statistically measured.

14.1 History and Meaning

We have already studied the concept of correlation in a previous chapter. Correlation presumes a bivariate distribution (Two variables, say X and Y , varying together). Historically, the idea of regression came first and the correlation method afterwards. Sir Francis Galton was studying the correlates of heredity with a view to verify Darwin's Theory of Evolution. He studied the heights of fathers and their sons. He began by preparing a scatter diagram, perhaps the first in the history of statistics. He converted all heights on a common scale, i.e., z scale, having a measuring unit of 1σ . He also computed the means of sons' heights in z scores for some fixed heights of parents. He noticed two important things:

1. The means of sons' heights fell along a straight-line trend, and
2. The means of sons' heights did not increase as rapidly as did the parents' heights. Each mean height of sons

deviated less from their general mean than the height of parents from which they came deviated from their own mean. He called this "falling back" of heights of sons toward the general mean as the law of filial regression. Regression thus implies "going back" or returning. Galton studied the average relationship between these two variables graphically and called the line describing this relationship, as the line of regression. Regression lines study the average relationship between two variables.

In a simple regression problem there are two regression lines—one for the regression of Y on X (predicting X from Y); and the other for the regression of X on Y (predicting Y from X). True, that if two variables are measured on each person in a group, the relationship should be the same regardless of whether one predicts Y from X or X from Y . However, the regression constants will be different depending upon which variable is the predictor and which is the predicted. However, if the correlation between X and Y is perfect and linear, the two regression lines will coincide. However since perfect relationships are rare in social sciences, there are usually two regression lines (See Figure 14.1).

14.2 Equation of a Straight Lines

A straight line can be described by its slope b ; and Y -intercept, a . Suppose a labourer charges Re. 1.00 per hour for his services. Let us label money earned as Y and hours worked as X . If the labourer works for 0 hours ($X=0$), he makes no money ($Y=0$); Working two hours, ($X=2$) fetches Rs. 2/—($Y=2$), and so on. The relationship can be shown as below:

X	Y
0	0
2	2
4	4
6	6

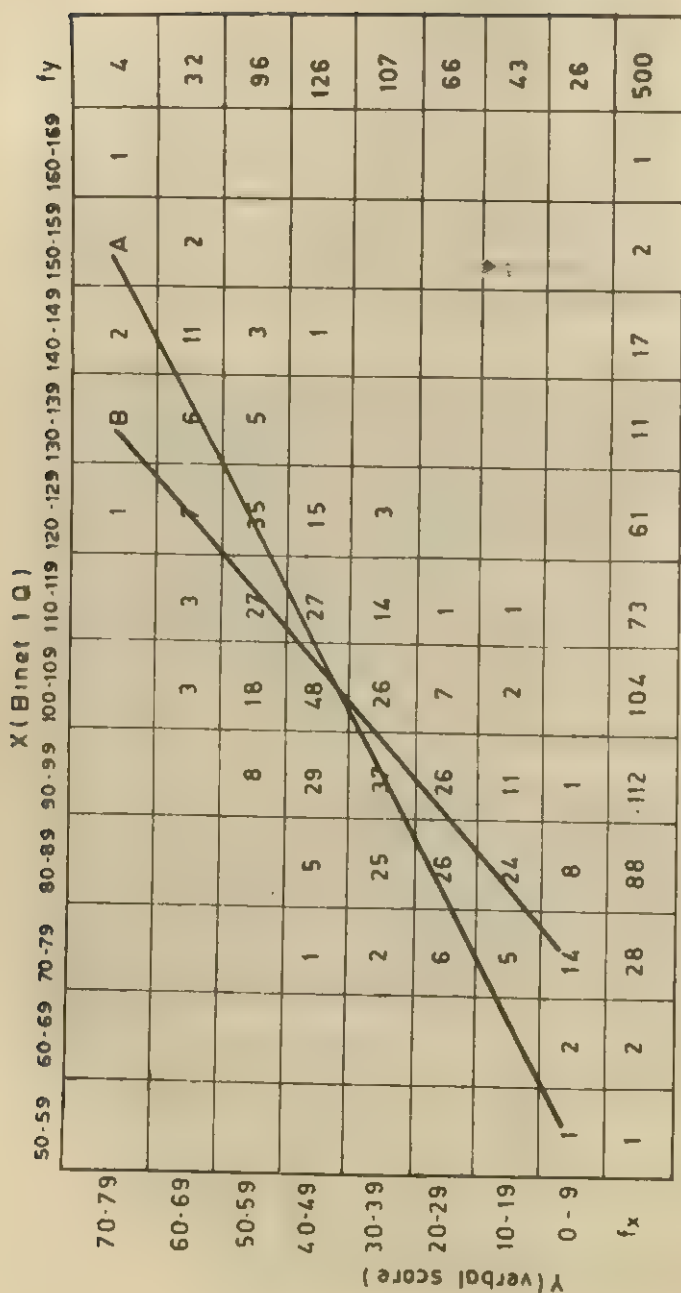


Fig. 14.1. Bivariate Frequency Distribution Showing Regressions Lines of Y on X , and X on Y .

This table is based on a relationship between the number of hours worked and the money earned. An hour worked is a rupee earned. This can be symbolically represented as follows:

$$Y=X$$

With the help of this simple mathematical equation, the amount of money earned can be calculated from any amount of time spent working. Equations for other relationships can also be set up similarly, as shown below.

$$y=.75x \text{ (An hour's work fetches Re. 0.75)}$$

$$y=.50x \text{ (An hour's work fetches Re. 0.50)}$$

.75 and .50 in these examples are the slopes respectively of the two straight lines. The slope of a straight line, b , coefficient, is defined as the ratio between the vertical distance and the horizontal distance between any two points on the line.

Suppose our labourer is paid Re. 1.0 as the registration amount in addition of his wages at the rate of Re. 0.75 per hour. Once the contract has been entered upon the payment of registration amount is obligatory even if the labourer is not called to work. The situation is illustrated in the table below:

X	Y
0	1.00
1	1.75
3	3.25
4	4.00

The equation for this situation would be as follow:

$$y=.75x+1.00$$

The constant of 1.00 added to the equation here is called Y -intercept a , or the height of the Y axis where the line intersects it. In the previous example, the value of a was zero,

as no constant was added to the equation. Hence a straight line can be described with the help of the following equation

$$Y = bx + a \quad (14.1)$$

This concept is basic to the setting up of regression lines for the sake of predicting one variable from the other.

14.3 Simple Regression

As mentioned earlier, for predicting a variable from another, it is essential to set up a regression equation, or equation of the regression line. Furthermore it has also been mentioned that the equation of a straight line is based upon two constants, a and b . Hence it is necessary to calculate the values of these constants before predictions can be made. The procedure of setting up regression equations will be demonstrated on raw data as well as on the information about coefficient of correlation, standard deviations and means.

14.4 Regression Equations from Raw Scores

Five subjects have been given two tests, X and Y . Their scores are given in Table 14.1 under the columns X and Y . Set up regression equations for predicting Y from X , and also X from Y .

TABLE 14.1

Simple Regression from Raw Scores

X	Y	X^2	Y^2	XY
10	12	100	144	120
11	18	121	324	198
12	20	144	400	240
9	10	81	100	90
8	10	64	100	80
50	70	510	1068	728
\bar{X}	\bar{Y}	\bar{X}^2	\bar{Y}^2	$\bar{X}\bar{Y}$
10	14	100	196	140
$N=5$				

Predicting Y from X Scores

Step I: Find the value of b coefficient for predicting Y from X.

$$b_{yx} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \quad (14.2)$$

Substituting the values

$$b_{yx} = \frac{(5)(728) - (50)(70)}{(5)(510) - (50)(50)} = 2.8$$

Step II: Find the value of a coefficient for predicting Y from X

The formula for calculating the value of a is as follows:

$$a_{yx} = M_y - b_{yx} M_x \quad (14.3)$$

Substituting the values, we obtain

$$\begin{aligned} a_{yx} &= 14 - (2.8)(10) \\ &= 14 - 28 \\ &= -14 \end{aligned}$$

Due care should be taken about the sign of this value, especially when it is negative.

Step III: Set up the full regression equation.

The full regression equation for predicting Y scores from X scores would be

$$Y' = bX + a \quad (14.4)$$

in which y' is the predicted y score and b and a are regression constants.

Substituting the values, we obtain

$$\begin{aligned} Y' &= 2.8X + (-14) \\ &= 2.8X - 14 \end{aligned}$$

It means that a Y score can be viewed as 2.8 times the X score with a constant of 14 subtracted.

Predicting X from Y scores

In this situation, Y is the independent variable and X is the criterion variable. The steps are the same as shown above for predicting Y from X. The formulas will, no doubt, undergo some change which may be noted carefully.

$$b_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum Y^2 - (\sum Y)^2} \quad (14.5)$$

There is a change in the denominator only.

Substituting the values, we obtain

$$b_{xy} = \frac{(5)(728) - (50)(70)}{5(1068) - (70)(70)} = .318$$

$$a_{xy} = M_x - b_{xy}M_y \quad (14.6)$$

Substituting the value, we obtain

$$\begin{aligned} a_{xy} &= 10 - (.318)(14) \\ &= 10 - 4.452 \\ &= 5.548 \end{aligned}$$

The regression equation for predicting X from Y will be:

$$\begin{aligned} X' &= bY + a \\ &= .318Y - 5.548 \end{aligned}$$

It means that an X score can be viewed as .318 times the Y score plus a constant of 5.548 added.

Values of regression coefficients a and b in both the situations can also be obtained by using and solving the simultaneous normal equations given below:

Predicting Y from X

$$\Sigma Y = Na + b \Sigma X \text{ and} \quad (14.8a)$$

$$\Sigma XY = a \Sigma X + b \Sigma X^2 \quad (14.8b)$$

Predicting X from Y

$$\Sigma X = Na + b \Sigma Y \text{ and} \quad (14.9a)$$

$$\Sigma XY = a \Sigma Y + b \Sigma Y^2 \quad (14.9b)$$

The student may try the above simultaneous normal equations to check the solution by the other method given earlier. The notation is the same as shown in Table 14.1.

14.5 Regression Equations from SD's, r and M's

In situations when standard deviations, coefficient of correlation and means are given it is advisable to use this information to set up regression equations for predicting Y from X, and X from Y.

Example: On the basis of the following information, set up the two regression equations:

	M	σ	r
X	65	6	.80
Y	75	8	

Predict values of Y for X=70, 72, 48 and 60.

Predicting Y from X

The values of the two regression coefficient a and b can be calculated as below:

$$b_{yx} = r_{yx} \left(\frac{\sigma_y}{\sigma_x} \right) \quad (14.10)$$

b_{yx} , with the subscripts in this order implies that we are predicting Y from X.

$$\begin{aligned} b_{yx} &= .8 \left(\frac{8}{6} \right) = 1.064 \\ a_{yx} &= M_y - (M_x) b_{yx} \\ &= 75 - (65) (1.064) \\ &= 5.84 \end{aligned} \quad (14.11)$$

The complete prediction equation will now be

$$Y = b_{yx}X + a_{yx}$$

$$Y = 1.064 + 5.84$$

The entire regression equation can also be obtained by using the composite formula given below

$$Y' = r \left(\frac{\sigma_y}{\sigma_x} \right) (X - M_x) + M_y \quad (14.12)$$

Substituting the values, we obtain

$$Y' = .8 (8/6) (X - 65) + 75$$

$$= 1.064 (X - 65) + 75$$

$$= 1.064X + 5.84 \text{ (checks with the previous result)}$$

Predicting X from Y

$$b_{xy} = r_{xy} \left(\frac{\sigma_x}{\sigma_y} \right) \quad (14.13)$$

$$= .8 (6/8)$$

$$= .6$$

$$a_{xy} = M_x - (M_y) b_{xy} \quad (14.14)$$

$$= 65 - (75) (.6)$$

$$= 20$$

The regression equation then will be

$$X = .6Y + 20$$

By using composite formula

$$X = r \left(\frac{\sigma_x}{\sigma_y} \right) (Y - M_y) + M_x \quad (14.15)$$

$$= (.8) (6/8) (Y - 75) - 65$$

$$= (.6) (Y - 75) - 65$$

$$= .6Y + 20 \text{ (Checks with previous result)}$$

On the basis of the regression equations set up by the process shown above, prediction of Y scores from X scores, and X scores from Y scores can be made simply by substituting the given values of a variable.

Predicting Y values from $X=70, 72, 48$ and 60

$$\therefore Y' = 1.064X + 5.84 \quad (\text{set up earlier})$$

$$\therefore X=70; Y' = (1.064)(70) + 5.84 = 80.32$$

$$X=72; Y' = (1.064)(72) + 5.84 = 82.45$$

$$X=48; Y' = (1.064)(48) + 5.84 = 56.85$$

$$X=60; Y' = (1.064)(60) + 5.84 = 69.68$$

Predicting X values from $Y=75, 65, 82$ and 72

$$X' = .6Y + 20 \quad (\text{Already set up})$$

Then

$$\text{for } Y=75; X' = (.6)(75) + 20 = 65.00$$

$$Y=65; X' = (.6)(65) + 20 = 59.00$$

$$Y=82; X' = (.6)(82) + 20 = 69.2$$

$$Y=72; X' = (.6)(72) + 20 = 67.2$$

The student may try some other values of one variable and predict the values of the other variable

14.6 Relationship between b coefficients and r

One important check of the accuracy of the two regression equations is that

$$b_{yx} b_{xy} = r^2 \quad (\text{Relation of } b\text{'s to } r^2) \quad (14.16)$$

$$\text{or } r = \sqrt{b_{yx} b_{xy}} \quad (14.17)$$

In other words the coefficient of correlation is equal to the square root of the product of the two b coefficients. In our example, $b_{yx} = 1.064$; and $b_{xy} = .6$

$$\text{Hence } r^2 = 1.064 \times .6$$

$$= .64 \text{ (rounded)}$$

$$r = \sqrt{.64}$$

$$= .8 \text{ (Checks with our } r \text{ above)}$$

Another check of the accuracy of the prediction equations is that the Y' for the M_x will be equal to M_y ; and X' for M_y will be equal to M_x .

14.7 Standard Error of the Estimates

The deviations between the predicted scores and the actual scores introduce an error. These deviations ($Y - Y'$ and $X - X'$) can be squared, summed, averaged and then the square root extracted. This index of the discrepancies between the observed and the predicted values is called the standard error of the estimates. When we predict on the basis of the regression equations, we need not calculate $Y - Y'$ or $X - X'$. The SE of the estimate can be calculated from the correlation coefficient and the standard deviation. The formulas are

SE of Estimate for predicting Y from X,

$$\sigma_{yx} = \sigma_y \sqrt{1 - r_{xy}^2} \quad (14.18)$$

SE of Estimate for predicting X from Y,

$$\text{and } \sigma_{xy} = \sigma_x \sqrt{1 - r_{xy}^2} \quad (14.19)$$

(Standard Errors of Estimate Computed from r)

In our example, using r and σ 's,

$$\sigma_{yx} = 8\sqrt{1 - .8} = 8 \times .447 = 3.576$$

and

$$\sigma_{xy} = 6\sqrt{1 - .8} = 6 \times .447 = 2.682$$

The interpretation of SE of the estimate is like that of SE of measurement. We assume normality and interpret the results by setting up probability or odds. In our example above an X score of 60 has a Y' score of 69.68. Hence the confidence interval for various odds can be set up as follows:

$$\begin{aligned} 68\% \text{ or } 2 \text{ to } 1 \text{ odds : } Y' \pm 1\sigma_{xy} & \quad (14.20) \\ & : 69.68 \pm 3.576 \\ \text{or : } 66.104 \text{ to } 73.256 \end{aligned}$$

Hence the odds are two to one that any individual whose X score is equal to 60 will not fall below 66.104 or go above 73.256, these scores being one σ_{yx} below and above the predicted Y .

$$\text{For 95\% : } Y' \pm 1.96 \sigma_{yx} \quad (14.21)$$

$$\text{For 99\% : } Y' \pm 2.58 \sigma_{yx} \quad (14.22)$$

The values of 1.96 and 2.58 have been taken from the normal curve area tables.

The student may try the interpretation of other scores at these levels.

14.8 Assumptions

The setting up of regression equation, making predictions, and calculation of the SE of the estimate involve certain assumptions which need to be made explicit.

Firstly, we assume that the relationship between the two variables is *linear*. It means that a straight line is the best way to describe the relationship. Linearity can be gauged by taking a look at the scatter plot. Statistical tests of linearity also exist.

Secondly, we assume *homoscedasticity* which means that Y scores in any single column have essentially the same standard deviation. Homoscedasticity can be translated as "similar variability" in each column.

Thirdly, we assume that Y scores within any one column are *normally distributed*. This assumption is essential when we interpret the predicted scores by using normal distribution.

14.9 Multiple Prediction

We have already discussed the correlation and regression based on two variables—one independent and the other dependent or criterion. It was a situation involving simple regression. However, actual relationships in Social Sciences are not always as simple as that. There could be two or more variables affecting or jointly related to a dependent variable. School marks may be a joint function of intelligence and the number of hours at study or a rural and urban setting. Hence

in such situations one has to keep in view multiple dependence instead of the idea of dependence of one variable on another single variable. Multiple dependence means the dependence of one variable on two or more independent variables. Success in sports may be related both to aptitude and training. Multiple dependence can be indicated by the statistic, the coefficient of multiple correlation, R .

14.10 The Coefficient of Multiple Correlation, R

Multiple correlation indicates the strength of relationship between dependent variable and two or more independent variables taken together. It is related to the inter-correlations among independent variables as well as to their correlations with the dependent variable. The process of calculation of the coefficient of multiple correlation, R is illustrated below by taking hypothetical data.

TABLE 14.2

Intercorrelation among Four Variables

<i>Variables</i>	X_2	X_3	X_4	X_1
X_2	—	.50	.60	.70
X_3	.50	—	.20	.80
X_4	.60	.20	—	.90
X_1	.70	.80	.90	—
M_x	73	55	60	78
σ_x	12	10	15	16

X_1 = Criterion or dependent variable

X_2, X_3, X_4 = Independent variables.

M_x and σ_x = Means and SD's respectively of the four variables.

Three-Variable Solution

We take a three-variable problem and demonstrate the process of calculation of R and setting up the regression

equation. The formula for the calculation of coefficient of multiple correlation is:

$$R^2_{1.23} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \quad (14.23)$$

$R_{1.23}$ is the square-root of $R^2_{1.23}$

r_{12} , r_{23} , r_{14} , r_{13} are correlations between pairs of variables as indicated by their subscripts.

$R_{1.23}$ = Coefficient of multiple correlation between X_1 and a combination of X_2 and X_3 .

Substituting the values of correlation coefficients in Formula (14.23), we obtain

$$\begin{aligned} R^2_{1.23} &= \frac{(.70)^2 + (.80)^2 - 2 \times .70 \times .80 \times .50}{1 - (.50)^2} \\ &= \frac{.49 + .64 - .56}{1 - .25} = .76 \\ &= R_{1.23} = \sqrt{.76} = .87 \end{aligned}$$

Formula 14.23 can be easily modified to obtain multiple correlation among other combinations of variables. For example

$$R^2_{1.34} = \frac{r_{13}^2 + r_{14}^2 - 2r_{13}r_{14}r_{34}}{1 - r_{34}^2} \quad (14.34)$$

The student may try to write the formula for $R^2_{1.24}$ himself.

We should remember two important principles regarding multiple correlation. These are based on the extent of correlations between the independent variables; and each independent variable with dependent variable.

1. An increase in the correlations between the dependent variable and independent variables leads to an increase in the value of R .

2. A decrease in the correlations between independent variables leads to an increase in the value of R .

From these principles, it is implied that a maximum R will be obtained when correlations with X_1 (criterion) are large and intercorrelations of X_2, X_3, \dots, X_m are small.

If intercorrelation of independent variables is zero, Formula (14.23) will be reduced to

$$R_{1.23}^2 = r_{12}^2 + r_{13}^2 \quad (14.25)$$

\therefore The last term in the numerator is reduced to 0; and the denominator to 1.

14.11 The Multiple-Regression Equation

In the preceding section, a multiple correlation has been calculated by taking a three variable problem. The same will be extended for the purpose of setting up the multiple regression equation for prediction of values of X_1 from knowledge of the values of X_2 and X_3 . The equation for multiple prediction for our problem will be:

$$X'_1 = a + b_{12.3} X_2 + b_{13.2} X_3 \quad (14.26)$$

In which, X'_1 = the predicted or dependent variable
 b coefficients = the multiplying constants or weights
 a coefficient = a constant to be added.

b Coefficients are the optimal weights which ensure maximization of correlation between predicted and observed X values. These are based on the principle of least squares.

To set up the prediction (14.26), we must solve for the value of b coefficients. The various formulas are:

$$b_{12.3} = \frac{\sigma_1}{\sigma_2} \beta_{12.3} \quad (14.27)$$

(Partial regression coefficient, keeping X_3 variable constant)

$$b_{13.2} = \frac{\sigma_1}{\sigma_3} \beta_{13.2} \quad (14.28)$$

(Partial regression coefficient, keeping X_2 variable constant)

in which σ_1 , σ_2 and σ_3 are standard deviations of variables X_1 , X_2 and X_3 . $\beta_{12.3}$ and $\beta_{13.2}$ are standard partial regression coefficients. The first partials out or keeps constant the effect of X_3 , while the second, that of X_2 , as is done in partial correlation. The betas, $\beta_{12.3}$ and $\beta_{13.2}$ are found as below:

$$\beta_{12.3} = \frac{r_{12} - r_{13} r_{23}}{1 - r_{23}^2} \quad (14.29)$$

$$\beta_{13.2} = \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \quad (14.30)$$

(Standard partial regression coefficients or β weights)

in which correlation coefficients, for various combinations of variables have been indicated by the subscripts.

Substituting the values from Table 14.2, we solve the formulas from 14.29 to 14.30. We first calculate the values of β weights:

$$\beta_{12.3} = \frac{.70 - (.80)(.50)}{1 - (.50)^2} = \frac{.70 - .40}{1 - .25} = 0.4$$

$$\beta_{13.2} = \frac{.80 - (.70)(.50)}{1 - (.50)^2} = \frac{.80 - .35}{1 - .25} = 0.6$$

We may now solve for the b coefficients by means of Formulas 14.27 and 14.28.

$$b_{12.3} = \left(\frac{16}{12} \right) (0.4) = 1.333 \times 0.4 = 0.533$$

$$b_{13.2} = \left(\frac{16}{10} \right) (0.6) = 1.6 \times 0.6 = 0.96$$

Now let us calculate the value of a constant

$$a_{1.23} = M_1 - b_{12.3} M_2 - b_{13.2} M_3 \quad (14.31)$$

in which M_1 , M_2 and M_3 are the means of the variables X_1 , X_2 and X_3 respectively:

Substituting the values, we obtain

$$\begin{aligned} a_{1.23} &= 78 - (0.4) (72) - (0.6) (55) \\ &= 78 - 28.8 - 33 \\ &= 16.2 \end{aligned}$$

Now substituting these values in equation (14.26)

$$\begin{aligned} X'_1 &= a + b_{12.3} X_2 + b_{13.2} X_3 \\ &= 16.2 + .533X_2 + .96X_3 \end{aligned}$$

Now we can predict the values of X_1 from the knowledge of the values of X_2 and X_3 .

For example, if $X_2=60$ and $X_3=40$, the predicted X'_1 score will be:

$$\begin{aligned} X'_1 &= 16.2 + (.533) (60) + (.96) (40) \\ &= 16.2 + 31.98 + 38.40 \\ &= 86.58 \end{aligned}$$

14.12 Calculation of R from Betas

Beta coefficients can also be used to calculate multiple R as follows:

$$\begin{aligned} R_{1.23}^2 &= \beta_{12.3} r_{12} + \beta_{13.2} r_{13} \\ &\quad (\text{R from Beta weights}) \end{aligned} \tag{14.32}$$

Substituting the values

$$\begin{aligned} R_{1.23}^2 &= (.4) (.7) + (.6) (.8) \\ &= .28 + .48 \\ &= .76 \\ R_{1.23} &= \sqrt{.76} = .87 \end{aligned}$$

(Checks with our value calculated earlier by Formula 14.23)

14.13 Standard Error of Estimate from Multiple Prediction

Standard error of estimate from multiple prediction can

be calculated to know as to how far the predicted values would deviate from the obtained ones. The formula for the purpose is:

$$\sigma_{1.23} = \sigma_1 \sqrt{1 - R_{1.23}^2} \quad (14.33)$$

in which, σ_1 is the SD of the X_1 variable, and

R^2 is the multiple R squared.

In our example,

$$\begin{aligned} \sigma_{1.23} &= 16\sqrt{1 - (.87)^2} \\ &= 16 \times .24 \\ &= 3.84 \end{aligned}$$

The interpretation of SE of estimate is similar to that in the case of simple regression. Here it can be said that two-thirds of the obtained X_1 values will lie between 3.84 points of the predicted X_1 values.

14.14 Other Methods

Multiple correlation with more than three variables can be calculated and prediction equations set up, by using the Doolittle-solution and other methods which are given in texts on advanced statistics. Statisticians have devised methods to find out a correlation between a combination of dependent variables and a combination of independent variables. It is called *Canonical correlation* and requires the help of a computer as the computation work involved is enormous. Multiple regression equations are used in the preparations of tests on the basis of the weights allotted to various sub-tests. The contribution of various components can also be calculated.

Exercises for Practice

- 14.1 Raw scores of five students on an intelligence test (X) and an academic achievement measure (Y) are given below. Set up regression equations for predicting Y from X; and X from Y.

<i>Persons</i>	1	2	3	4	5
X	5	3	2	8	2
Y	7	5	2	10	6

- 14.2 (a) From the following information, set up regression equations for predicting Y from X; and X from Y.

	M	σ	r
X	130	10	.70
Y	110	18	

- (b) Predict values of X for Y = 120, 90, 125 and 130
 (c) Predict values of Y for X = 122, 115, 140, 135, 132 and 128.
- 14.3 Compute the regression equations for the prediction of Y, from the following set of data:

X	Y
0	1
1	3
2	2
3	4
4	5

- 14.4 What do you mean by,

- (a) Regression (b) Standard error of estimate (c) Multiple correlation (d) b and β coefficients.

- 14.5 (a) From Table 14.2 set up the multiple regression equation for predicting X_1 from X_2 and X_3 . Calculate SE of the estimate.

- (b) Predict X_1 from: $X_2 = 45$, $X_3 = 65$; $X_2 = 50$, $X_3 = 62$; $X_2 = 40$, $X_3 = 70$.

CHAPTER 15

AN INTRODUCTORY NOTE ON SECOND GENERATION OF MULTIVARIATE ANALYSIS*

(A potential improvement in the methodology for social research)

The growing methodological recognition that scientific theory involves both abstract and empirical variables has resulted in the creation of a new class of multivariate data analysis techniques. Here the objective is to bring data and theory together. A sound theory is supposed to cover aspects of generality, integration of concepts and parsimony, and also, at the same time should be substantiated by concrete data so as to make it realistic and free from meaningless abstractions and imaginations. The second generation methods attempt at achieving this important goal.

15.1 Distinguishing Characteristics

The criteria on the basis of which the second generation methods can be distinguished from the first generation methods are:

- (1) The analysis of the nature of theoretical constructs in the model.
 - (2) Incorporation of the nature of construct relationships in the model, and also
 - (3) The conceptualisation of epistemic relationships in the model.
- Multi-dimensional approach to observing and multivariate approach to data analysis is a logical necessity in social sciences and while realizing this, any sound thinker is bound to also realize that dependency on an analysis of pure empirical variables can also lead to serious errors of inference and misinterpretations especially in social sciences. The second generation methods attempt to fill up this gap and suggest improvements for making theories more realistic.

* Dr P. L. NARANKH Research Associate (UGC), South Gujarat University actively helped in preparing this note.

Examples of multivariate first generation methods are: factor analysis, cluster analysis, principal components analysis, discriminant analysis. These techniques required fewer assumptions and less apriori theoretical knowledge and hence ultimately could not be applied in the right perspective. The second generation methods is an improvement in these weaker aspects and thus obviously requires more theoretical assumptions and incorporate more apriori information in the model. An additional justification and substantiation of these theoretical issues is of course again a logical necessity.

15.2 Second Generation Methods: An obvious extension of First Generation Techniques

A look at the list of the following examples of second generation methods makes it obvious that each is an extension of some or other first generation techniques. This extension is just because of incorporation of apriori theoretical issues into the overall considerations.

- (1) Redundancy Analysis: An improvisation over canonical analysis.
- (2) External single set components analysis. It is also another improvement over the classical canonical correlation analysis and;
- (3) Analysis of Linear Structural Relationships (LISREL Model) : An improvement over the linear model.
- (4) Factor Analytic Structural Equations Model: An improvement over the factor analytic techniques.
- (5) The Partial Least Squares Components Structural Equation Model: An improved version of least squares methodology; and
- (6) The constrained confirmatory Monotone Distance Analysis: An improvisation of the multi-dimensional scaling model.

15.3 Issues of Variables and their relationships in the Second Generation Methods

Usage of the word "construct" instead of "variable" is quite common in more advanced situations. A construct is a variable that is of interest to the substantive context under examination. A defined construct is a composite of its indicators and always an estimate in the equation instead of the true value. Estimation is generally the result of minimization of error in measurement. In the theoretical considerations two constructs may depict the relationships of orthogonality (zero correlation); symmetry (no distinction in the direction of relationships)

or directionality. These relationships are preconceptualized in the second generation methods and then based on these conceptualisations equations or models are derived to describe the situations or to derive the tests of hypotheses. The conceptualisations about epistemic relationships describe the link between theory and data and generally comprise of the rules of correspondence or the so called correspondence postulates.

15.4 Concluding Remarks

For want of space it is difficult to illustrate the full aspects of these newly emerging techniques, but it may be realized that these methods which are gaining more and more in popularity are out of theoretical necessities in social research and have arisen out of the realization that an analysis of pure empirical variables can never describe reality or throw sufficient light on the hypotheses to be tested. They are definitely a potential improvement in the methodology of social research in the coming years.

For Further Reading

1. Fornell C. (Ed.): *A Second Generation of Multivariate Analysis*, Vols. I & II, New York: PRAEGER; 1982.
2. Bohrnstedt G.W. & Borgatta F.E. (Eds.), *Social Measurement: Current Issues*, Beverly Hills, SAGE; 1981.
3. Shye S. (Eds.), *Theory Construction and Data Analysis in the Behavioural Sciences*, San Francisco: Jossey Bass; 1978.

APPENDICES

<i>Table</i>	<i>Description</i>
A	Proportions of Area under Normal Distribution Curve.
B	Critical Values of t
C	Conversion of r into z
D	Critical Values of Pearson Product Moment Correlation.
E	Critical Values of Spearman Rank Correlation.
F	Critical Values of Chi-square.
G	Probabilities Associated with Values as Small as observed Values of x in the Binomial Test.
H	Significance of a_3 (Skewness)
I	Significance of a_4 (Kurtosis)
J	Critical Values of K in the Kolmogorov—Smirnov Two Sample Test when Samples are small.
K	F-Ratio
L	A Table to Aid in the Calculation of T Scores
M	Standard Scores (or deviates) and ordinates Corresponding to divisions of the area under the Normal Curve into a larger portion (B) and a smaller portion (C); also the Value \sqrt{BC} .
N	Values of r , taken as the cosine of an angle
O	Table of Squares and Square Roots of the Numbers From 1—1000

TABLE A

Fractional parts of the total area (taken as 0.1000) under the normal probability curve, corresponding to distances on the base-line between the mean and successive points laid off from the mean in units of standard deviation

Example: Between the mean and a point 1.38σ ($\frac{x}{\sigma} = 1.38$) are found 41.62% of the entire area under the curve.

$\frac{x}{\sigma}$.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549

0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952

TABLE B

Table of t , for use in determining the significance of statistics

Example: When the df are 35 and $t=2.03$, the .05 in column 3 means that 5 times in 100 trials a divergence as large as that obtained may be expected in the positive *and* negative directions under the null hypothesis.

Degrees of Freedom	Probability (P)			
	0.10	0.05	0.02	0.01
(1)	(2)	(3)	(4)	(5)
1	$t=6.34$	$t=12.71$	$t=31.82$	$t=63.66$
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.90	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.06
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86

(1)	(2)	(3)	(4)	(5)
20	1.72	2.09	2.53	2.84
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.50	2.81
24	1.71	2.06	2.49	2.80
25	1.71	2.06	2.48	2.79
26	1.71	2.06	2.48	2.78
27	1.70	2.05	2.47	2.77
28	1.70	2.05	2.47	2.76
29	1.70	2.04	2.46	2.76
30	1.70	2.04	2.46	2.75
35	1.69	2.03	2.44	2.72
40	1.68	2.02	2.42	2.71
45	1.68	2.02	2.41	2.69
50	1.68	2.01	2.40	2.68
60	1.67	2.00	2.39	2.66
70	1.67	2.00	2.38	2.65
80	1.66	1.99	2.38	2.64
90	1.66	1.99	2.37	2.63
100	1.66	1.98	2.36	2.63
125	1.66	1.98	2.36	2.62
150	1.66	1.98	2.35	2.61
200	1.65	1.97	2.35	2.60
300	1.65	1.97	2.34	2.59
400	1.65	1.97	2.34	2.59
500	1.65	1.96	2.33	2.59
1000	1.65	1.96	2.33	2.58
∞	1.65	1.96	2.33	2.58

TABLE C

Conversion of a Pearson r into a corresponding
Fisher's z coefficient*

r	z	r	z	r	z	r	z	r	z	r	z
.25	.26	.40	.42	.55	.62	.70	.87	.85	1.26	.950	1.83
.26	.27	.41	.44	.56	.63	.71	.89	.86	1.29	.955	1.89
.27	.28	.42	.45	.57	.65	.72	.91	.87	1.33	.960	1.95
.28	.29	.43	.46	.58	.66	.73	.93	.88	1.38	.965	2.01
.29	.30	.44	.47	.59	.68	.74	.95	.89	1.42	.970	2.09
.30	.31	.45	.48	.60	.69	.75	.97	.90	1.47	.975	2.18
.31	.32	.46	.50	.61	.71	.76	1.00	.905	1.50	.980	2.30
.32	.33	.47	.51	.62	.73	.77	1.02	.910	1.53	.985	2.44
.33	.34	.48	.52	.63	.74	.78	1.05	.915	1.56	.990	2.65
.34	.35	.49	.54	.64	.76	.79	1.07	.920	1.59	.995	2.99
.35	.37	.50	.55	.65	.78	.80	1.10	.925	1.62		
.36	.38	.51	.56	.66	.79	.81	1.13	.930	1.66		
.37	.39	.52	.58	.67	.81	.82	1.16	.935	1.70		
.38	.40	.53	.59	.68	.83	.83	1.19	.940	1.74		
.39	.41	.54	.60	.69	.85	.84	1.22	.945	1.78		

* r 's under .25 may be taken as equivalent to z 's

TABLE D

Correlation coefficients at the 5° and 1° levels of significance

Example: When N is 52 and df is 50, an r must be .273 to be significant at .05 level, and .354 to be significant at .01 level.

<i>Degrees of freedom ($N-2$)</i>	<i>.05</i>	<i>.01</i>	<i>Degrees of freedom ($N-2$)</i>	<i>.05</i>	<i>.01</i>
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.233	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

TABLE E

Values of rank-difference coefficients of correlation
that are significant at the .05 and .01 levels
(one-tail test)*

<i>N</i>	.05	.01	<i>N</i>	.05	.01
5	.900	1.000	16	.425	.601
6	.829	.943	18	.399	.564
7	.714	.893	20	.377	.534
8	.643	.833	22	.359	.508
9	.600	.783	24	.343	.485
10	.564	.746	26	.329	.465
12	.506	.712	28	.317	.448
14	.456	.645	30	.306	.432

*For a two-tail test, double the probabilities to .01 and .02.

TABLE F

χ^2 Table. P gives the probability of exceeding the tabulated value of χ^2 for the specified number of degrees of freedom (df). The values of χ^2 are printed in the body of the table*

df	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.00393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217

13	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

TABLE G

Cumulative proportions from the tail categories of binomial distribution for $(\frac{1}{2} + \frac{1}{2})^N$, with
 N varying from 6 to 25

Categories C

C N	$C\ 0$ N	$C\ 1$ $(N-1)$	$C\ 2$ $(N-2)$	$C\ 3$ $(N-3)$	$C\ 4$ $(N-4)$	$C\ 5$ $(N-5)$	$C\ 6$ $(N-6)$	$C\ 7$ $(N-7)$	$C\ 8$ $(N-8)$	$C\ 9$ $(N-9)$
6	.016	.109								
7	.008	.062	.227							
8	.004	.035	.145							
9	.002	.020	.090	.254						
10	.001	.011	.055	.172						
11		.006	.033	.113						
12		.003	.019	.073	.194					
13		.002	.011	.046	.133					

14	.001	.006	.029	.090			
15		.004	.018	.059	.151		
16		.002	.011	.038	.105		
17		.001	.006	.025	.072	.166	
18		.001	.004	.015	.048	.119	
19			.002	.010	.032	.084	.180
20			.001	.006	.021	.058	.132
21			.001	.004	.013	.039	.095
22				.002	.008	.026	.067
23				.001	.005	.017	.047
24				.001	.003	.011	.032
25					.002	.007	.022
							.054
							.154
							.115

COMMENT Each entry is the probability of an outcome as extreme as the last category (0 heads or N heads, as in coin tossing), the next-to-the-last category (1 head or $N-1$ heads), and so on (i.e., the probabilities are cumulative). Each probability is for one tail only. For a two-tail test, double the probability given.

TABLE H

Upper 0.10 and 0.02 limits of a_3 (Skewness) when computed from random samples from a normal population

<i>N</i>	0.10	0.02
50	.285	.619
75	.198	.424
100	.152	.321
125	.123	.258
150	.103	.216
175	.089	.185
200	.078	.162
250	.063	.130
300	.053	.108
350	.045	.093
400	.040	.081
450	.035	.072
500	.032	.065
550	.029	.059
600	.027	.054
650	.025	.050
700	.023	.046
750	.021	.043
800	.030	.041
850	.019	.038
900	.018	.036
950	.017	.034
1000	.016	.032

(Contd.)

TABLE H (Continued)

<i>N</i>	0.10	0.02
1200	.013	.027
1400	.012	.023
1600	.010	.020
1800	.009	.018
2000	.008	.016
2500	.006	.013
3000	.005	.011
3500	.005	.009
4000	.004	.008
4500	.004	.007
5000	.003	.006

TABLE I

Upper and Lower 0.05 and 0.01 limits of a_1 (Kurtosis)
when computed from random samples from a
normal population

<i>N</i>	<i>Lower limits</i>		<i>Upper limits</i>	
	0.01	0.05	0.05	0.01
100	2.18	2.35	3.77	4.39
125	2.24	2.40	3.71	4.24
150	2.29	2.45	3.65	4.13
175	2.33	2.48	3.61	4.05

(Contd.)

TABLE 1 (*Continued*)

<i>N</i>	<i>Lower limits</i>		<i>Upper limits</i>	
	<i>0.01</i>	<i>0.05</i>	<i>0.05</i>	<i>0.01</i>
200	2.37	2.51	3.57	3.98
250	2.42	2.55	3.52	3.87
300	2.46	2.59	3.47	3.79
350	2.50	2.62	3.44	3.72
400	2.52	2.64	3.41	3.67
450	2.55	2.66	3.39	3.63
500	2.57	2.67	3.37	3.60
550	2.58	2.69	3.35	3.57
600	2.60	2.70	3.34	3.54
650	2.61	2.71	3.33	3.52
700	2.62	2.72	3.31	3.50
750	2.64	2.73	3.30	3.48
800	2.65	2.74	3.29	3.46
850	2.66	2.74	3.28	3.45
900	2.66	2.75	3.28	3.43
950	2.67	2.76	3.27	3.42
1000	2.68	2.76	3.36	3.41
1200	2.71	2.78	3.24	3.37
1400	2.72	2.80	3.22	3.34
1600	2.74	2.81	3.21	3.32
1800	2.76	2.82	3.20	3.30
2000	2.77	2.83	3.18	3.28
2500	2.79	2.85	3.16	3.25
3000	2.81	2.86	3.15	3.22
3500	2.82	2.87	3.14	3.21
4000	2.83	2.88	3.13	3.19
4500	2.84	2.88	3.12	3.18
5000	2.85	2.89	3.12	3.17

TABLE J

Critical values of K in the Koimogrov-Smirnov two-sample test, when samples are small

N	One-tail test		Two-tail test		N	One-tail test		Two-tail test	
	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$		$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
3	—	—	—	—	18	8	10	9	10
4	—	—	4	—	19	8	10	9	10
5	5	5	5	5	20	8	10	9	11
6	6	6	5	6	21	8	10	9	11
7	6	6	6	6	22	9	11	9	11
8	6	6	6	6	23	9	11	10	11
9	7	7	6	7	24	9	11	10	12
10	7	7	7	7	25	9	11	10	12
11	8	8	7	8	26	9	11	10	12
12	8	8	7	8	27	9	2	10	12
13	8	8	7	8	28	10	12	10	12
14	8	8	8	8	29	10	12	11	13
15	9	9	8	8	30	10	12	11	13
16	9	9	8	8	35	11	13	11	13
17	9	9	8	8	40	11	14	12	13

TABLE K
F-ratios for .05 (roman) and .01 (Italics) levels of significance

	<i>Degrees of freedom for greater mean square</i>									
	1	2	3	4	5	6	8	12	24	∞
1	161.45 4052.10	199.50 4999.03	215.72 5403.49	224.57 5625.14	230.17 5764.08	233.97 5859.39	238.89 5981.34	243.91 6105.83	249.04 6230.16	254.32 6366.48
2	18.51 98.49	19.00 99.01	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.37 99.36	19.41 99.42	19.45 99.46	19.50 99.50
3	10.13 34.12	9.55 30.81	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.84 27.49	8.74 27.05	8.64 26.60	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.04 14.80	5.91 14.37	5.77 13.93	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.82 10.27	4.68 9.89	4.53 9.47	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.15 8.10	4.00 7.72	3.84 7.31	3.67 6.88
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.73 6.84	3.57 6.47	3.41 6.07	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.44 6.03	3.28 5.67	3.12 5.28	2.93 4.86

9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.23 5.47	3.07 5.11	2.90 4.73	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.07 5.06	2.91 4.71	2.74 4.33	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	2.95 4.74	2.79 4.40	2.61 4.02	2.40 3.60
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.85 4.50	2.69 4.16	2.50 3.78	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.77 4.30	2.60 3.96	2.42 3.59	2.21 3.16
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.70 4.14	2.53 3.80	2.35 3.43	2.16 3.00
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.64 4.00	2.48 3.67	2.29 3.29	2.07 2.87
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.59 3.89	2.42 3.55	2.24 3.18	2.01 2.75
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.55 3.79	2.38 3.45	2.19 3.08	1.96 2.65
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.51 3.71	2.34 3.37	2.15 3.01	1.92 2.57

TABLE K (Continued)

		Degrees of freedom for greater mean square									
		1	2	3	4	5	6	8	12	24	∞
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88	
	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84	
	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81	
	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78	
	7.94	5.72	4.82	4.31	3.99	3.75	3.45	3.12	2.75	2.30	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76	
	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73	
	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71	
	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69	
	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67	
	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10	

28	4.20	3.34	2.93	2.71	2.56	2.44	2.29	2.12	1.91	1.65
	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
	7.60	5.42	4.34	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
35	4.12	3.26	2.87	2.64	2.48	2.37	2.22	2.04	1.83	1.57
	7.42	5.27	4.40	3.91	3.59	3.37	3.07	2.74	2.37	1.90
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.52
	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.82
45	4.06	3.21	2.81	2.58	2.42	2.31	2.15	1.97	1.76	1.48
	7.23	5.11	4.25	3.77	3.45	3.23	2.94	2.61	2.23	1.75
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	1.95	1.74	1.44
	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.56	2.18	1.68
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
70	3.98	3.13	2.74	2.50	2.35	2.23	2.07	1.89	1.67	1.35
	7.01	4.92	4.07	3.60	3.29	3.07	2.78	2.45	2.07	1.53
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.88	1.65	1.31
	6.96	4.88	4.04	3.56	3.26	3.04	2.74	2.42	2.03	1.47
90		3.10	2.71	2.47	2.32	2.20	2.04	1.86	1.64	1.28
	6.92	4.85	4.01	3.53	3.23	3.01	2.72	2.39	2.00	1.43

TABLE K (Continued)

Degrees of freedom for greater mean square										
	1	2	3	4	5	6	8	12	28	∞
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.21	2.19 2.99	2.03 2.69	1.85 2.37	1.63 1.98	1.26 1.39
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.01 2.66	1.83 2.33	1.60 1.94	1.21 1.32
150	3.90 6.81	3.06 4.75	2.66 3.91	2.43 3.45	2.27 3.14	2.16 2.92	2.00 2.63	1.82 2.31	1.59 1.92	1.18 1.27
200	3.89 6.76	3.04 4.71	2.65 3.88	2.42 3.41	2.26 3.11	2.14 2.89	1.98 2.60	1.80 2.28	1.57 1.88	1.14 1.21
300	3.87 6.72	3.03 4.68	2.64 3.85	2.41 3.38	2.25 3.08	2.13 2.86	1.97 2.57	1.79 2.24	1.55 1.85	1.10 1.14
400	3.86 6.70	3.02 4.66	2.63 3.83	2.40 3.37	2.24 3.06	2.12 2.85	1.96 2.56	1.78 2.23	1.54 1.84	1.07 1.11
500	3.86 6.69	3.01 4.65	2.62 3.82	2.39 3.36	2.23 3.05	2.11 2.84	1.96 2.55	1.77 2.22	1.54 1.83	1.06 1.08
1000	3.85 6.66	3.00 4.63	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	1.95 2.53	1.76 2.20	1.53 1.81	1.03 1.04
∞	3.84 6.64	2.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	1.94 2.51	1.75 2.18	1.52 1.79	

TABLE L

A Table to aid in the calculation of T scores

<i>Proportion below the point</i>	<i>T score</i>	<i>Proportion below the point</i>	<i>T score</i>	<i>Proportion below the point</i>	<i>T score</i>
.0005	17.1	.100	37.2	.900	62.8
.0007	18.1	.120	38.3	.910	63.4
.0010	19.1	.140	39.2	.920	64.1
.0015	20.3	.160	40.1	.930	64.8
.0020	21.2	.180	40.8	.940	65.5
.0025	21.9	.200	41.6	.950	66.4
.0030	22.5	.220	42.3	.960	67.5
.0040	23.5	.250	43.3	.965	68.1
.0050	24.2	.300	44.8	.970	68.8
.0070	25.4	.350	46.1	.975	69.6
.010	26.7	.400	47.5	.980	70.5
.015	28.3	.450	48.7	.985	71.7
.020	29.5	.500	50.0	.990	73.3
.025	30.4	.550	51.3	.993	74.6
.030	31.2	.600	52.5	.995	75.8
.035	31.9	.650	53.9	.9960	76.5
.040	32.5	.700	55.2	.9970	77.5
.050	33.6	.750	56.7	.9975	78.1
.060	34.5	.780	57.7	.9980	78.7
.070	35.2	.800	58.4	.9985	79.7
.080	35.9	.820	59.2	.9990	80.9
.090	36.6	.840	59.9	.9993	81.9
		.860	60.8	.9995	82.9
		.880	61.7		

TABLE M

Standard scores (or deviates) and ordinates corresponding to divisions of the area under the normal curve into a larger proportion (B) and a smaller proportion (C); also the value \sqrt{BC} .

<i>B</i> <i>The Larger</i> <i>area</i>	<i>Z</i> <i>Standard</i> <i>score</i>	<i>Y</i> <i>Ordinate</i>	\sqrt{BC}	<i>C</i> <i>The smaller</i> <i>area</i>
1	2	3	4	5
.500	.0000	.3989	.5000	.500
.505	.0125	.3989	.5000	.495
.510	.0251	.3988	.4999	.490
.515	.0376	.3987	.4998	.485
.520	.0502	.3984	.4996	.480
.525	.0627	.3982	.4994	.475
.530	.0753	.3978	.4991	.470
.535	.0878	.3974	.4988	.465
.540	.1004	.3969	.4984	.460
.545	.1130	.3964	.4980	.455
.550	.1257	.3958	.4975	.450
.555	.1383	.3951	.4970	.445
.560	.1510	.3944	.4964	.440
.565	.1637	.3936	.4958	.435
.570	.1764	.3928	.4951	.430
.575	.1891	.3919	.4943	.425
.580	.2019	.3909	.4936	.420
.585	.2147	.3899	.4927	.415
.590	.2275	.3887	.4918	.410
.595	.2404	.3876	.4909	.405

1	2	3	4	5
.600	.2533	.3863	.4899	.400
.605	.2663	.3850	.4889	.395
.610	.2793	.3837	.4877	.390
.615	.2924	.3822	.4867	.385
.620	.3055	.3808	.4854	.380
.625	.3186	.3792	.4841	.375
.630	.3319	.3776	.4828	.370
.635	.3451	.3759	.4814	.365
.640	.3585	.3741	.4800	.360
.645	.3719	.3723	.4785	.355
.650	.3853	.3704	.4770	.350
.655	.3989	.3684	.4754	.345
.660	.4125	.3664	.4737	.340
.665	.4261	.3643	.4720	.335
.670	.4399	.3621	.4702	.330
.675	.4538	.3599	.4684	.325
.680	.4677	.3576	.4665	.320
.685	.4817	.3552	.4645	.315
.690	.4959	.3528	.4625	.310
.695	.5101	.3503	.4604	.305
.700	.5244	.3477	.4583	.300
.705	.5388	.3450	.4560	.295
.710	.5534	.3423	.4538	.290
.715	.5681	.3395	.4514	.285
.720	.5828	.3366	.4490	.280
.725	.5978	.3337	.4465	.275
.730	.6128	.3306	.4440	.270
.735	.6280	.3275	.4413	.265
.740	.6433	.3244	.4386	.260
.745	.6588	.3211	.4359	.255

TABLE M (Continued)

1	2	3	4	5
.750	.6745	.3178	.4330	.250
.755	.6903	.3144	.4301	.245
.760	.7063	.3109	.4271	.240
.765	.7225	.3073	.4240	.235
.770	.7388	.3036	.4208	.230
.775	.7554	.2999	.4176	.225
.780	.7722	.2961	.4142	.220
.785	.7892	.2922	.4108	.215
.790	.8064	.2882	.4073	.210
.795	.8239	.2841	.4037	.205
.800	.8416	.2800	.4000	.200
.805	.8596	.2757	.3962	.195
.810	.8779	.2714	.3923	.190
.815	.8965	.2669	.3883	.185
.820	.9154	.2624	.3842	.180
.825	.9346	.2578	.3800	.175
.830	.9542	.2531	.3756	.170
.835	.9741	.2482	.3712	.165
.840	.9945	.2433	.3666	.160
.845	1.0152	.2383	.3619	.155
.850	1.0364	.2332	.3571	.150
.855	1.0581	.2279	.3521	.145
.860	1.0803	.2226	.3470	.140
.865	1.1031	.2171	.3417	.135
.870	1.1264	.2115	.3363	.130
.875	1.1503	.2059	.3307	.125
.880	1.1750	.2000	.3250	.120
.885	1.2004	.1941	.3190	.115

1	2	3	4	5
.890	1.2265	.1880	.3129	.110
.895	.2536	.1818	.3066	.105
.900	1.2816	.1755	.3000	.100
.905	1.3106	.1690	.2932	.095
.910	1.3408	.1624	.2862	.090
.915	1.3722	.1556	.2789	.085
.920	1.4051	.1487	.2713	.080
.925	1.4395	.1416	.2634	.075
.930	1.4757	.1343	.2551	.070
.935	1.5141	.1268	.2465	.065
.940	1.5548	.1191	.2375	.060
.945	1.5982	.1112	.2280	.055
.950	1.6449	.1031	.2179	.050
.955	1.6954	.0948	.2073	.045
.960	1.7507	.0862	.1960	.040
.965	1.8119	.0773	.1838	.035
.970	1.8808	.0680	.1706	.030
.975	1.9600	.0584	.1561	.025
.980	2.0537	.0484	.1410	.020
.985	2.1701	.0379	.1226	.015
.990	2.3263	.0267	.0995	.010
.995	2.5758	.0145	.0705	.005
.996	2.6521	.0118	.0631	.004
.997	2.7478	.0091	.0547	.003
.998	2.8782	.0063	.0447	.002
.999	3.0902	.0034	.0316	.001
.9995	3.2905	.0018	.0224	.0005

TABLE N

Volume of r , taken as the cosine of an angle

Example: Suppose that $r_1 = \cos 45^\circ$. Then $\cos 45^\circ = .707$, and $r_1 = .71$ (to two decimals)

Angle	Cosine	Angle	Cosine	Angle	Cosine
0	1.000	41	.755	73	.292
		42	.743	74	.276
5	.996	43	.731	75	.259
		44	.719	76	.242
10	.985	45	.707	77	.225
		46	.695	78	.208
		47	.682	79	.191
15	.966	48	.669	80	.174
16	.961	49	.656		
17	.956	50	.643	81	.156
18	.951			82	.139
19	.946	51	.629	83	.122
20	.940	52	.616	84	.105
		53	.602	85	.087
21	.934	54	.588		
22	.927	55	.574		
23	.921	56	.559	90	.000
24	.914	57	.545		
25	.906	58	.530		
26	.899	59	.515		
27	.891	60	.500		
28	.883				
29	.875	61	.485		
30	.866	62	.469		
		63	.454		
31	.857	64	.438		
32	.848	65	.423		
33	.839	66	.407		
34	.829	67	.391		
35	.819	68	.375		
36	.809	69	.358		
37	.799	70	.342		
38	.788				
39	.777	71	.326		
40	.766	72	.309		

TABLE O

Squares and Square Roots of the number from
1 to 1000

<i>Number</i>	<i>Square</i>	<i>Square Root</i>	<i>Number</i>	<i>Square</i>	<i>Square Root</i>
1	2	3	4	5	6
1	1	1.000	2	4	1.414
3	9	1.732	4	16	2.000
5	25	2.236	6	36	2.449
7	49	2.646	8	64	2.828
9	81	3.000	10	1 00	3.162
11	1 21	3.317	12	1 44	3.464
13	1 69	3.606	14	1 96	3.742
15	2 25	3.873	16	2 56	4.000
17	2 89	4.123	18	3 24	4.243
19	3 61	4.359	20	4 00	4.472
21	4 41	4.583	22	4 84	4.690
23	5 29	4.796	24	5 76	4.899
25	6 25	5.000	26	6 76	5.099
27	7 29	5.196	28	7 84	5.292
29	8 41	5.385	30	9 00	5.477
31	9 61	5.568	32	10 24	5.657
33	10 89	5.745	34	11 56	5.831
35	12 25	5.916	36	12 96	6.000
37	13 69	6.083	38	14 44	6.164
39	15 21	6.245	40	16 00	6.325
41	16 81	6.403	42	17 64	6.481
43	18 49	6.557	44	19 36	6.633
45	20 25	6.708	46	21 16	6.782
47	22 09	6.856	48	23 04	6.928
49	24 01	7.000	50	25 00	7.071
51	26 01	7.141	52	27 04	7.211
53	28 09	7.280	54	29 16	7.348
55	30 25	7.416	56	31 36	7.483
57	32 49	7.550	58	33 64	7.616
59	34 81	7.681	60	36 00	7.746

TABLE O (Contd.)

1	2	3	4	5	6
61	37 21	7.810	62	38 44	7.874
63	39 69	7.937	64	40 96	8.000
65	42 25	8.062	66	43 56	8.124
67	44 89	8.185	68	46 24	8.246
69	47 61	8.307	70	49 00	8.367
71	50 41	8.428	72	51 84	8.485
73	53 29	8.544	74	54 76	8.602
75	56 25	8.660	76	57 76	8.718
77	59 29	8.775	78	60 84	8.832
79	62 41	8.888	80	64 00	8.944
81	65 61	9.000	82	67 24	9.055
83	68 89	9.110	84	70 56	9.165
85	72 25	9.220	86	73 96	9.274
87	75 69	9.327	88	77 44	9.381
89	79 21	9.434	90	81 00	9.487
91	82 81	9.539	92	84 64	9.592
93	86 49	9.644	94	88 36	9.695
95	90 25	9.747	96	92 16	9.798
97	94 09	9.849	98	96 04	9.899
99	98 01	9.950	100	1 00 00	10.000
101	1 02 01	10.050	102	1 04 04	10.100
103	1 06 09	10.149	104	1 08 16	10.193
105	1 10 25	10.247	106	1 12 36	10.296
107	1 14 49	10.344	108	1 16 64	10.392
109	1 18 81	10.440	110	1 21 00	10.488
111	1 23 21	10.536	112	1 25 44	10.583
113	1 27 69	10.630	114	1 29 96	10.677
115	1 32 25	10.724	116	1 34 56	10.770
117	1 36 89	10.817	118	1 39 24	10.863
119	1 41 61	10.909	120	1 44 00	10.954
121	1 46 41	11.000	122	1 48 84	11.045
123	1 51 29	11.091	124	1 53 76	11.136
125	1 56 25	11.180	126	1 58 76	11.225
127	1 61 29	11.269	128	1 63 84	11.314

1	2	3	4	5	6
129	1 66 41	11.358	130	1 69 00	11.402
131	1 71 61	11.446	132	1 74 24	11.489
133	1 76 89	11.533	134	1 79 56	11.576
135	1 82 25	11.619	136	1 84 96	11.662
137	1 87 69	11.705	138	1 90 44	11.747
139	1 93 21	11.790	140	1 96 00	11.832
141	1 98 81	11.874	142	2 01 64	11.916
143	2 04 49	11.958	144	2 07 36	12.000
145	2 10 25	12.042	146	2 13 16	12.083
147	2 16 09	12.124	148	2 19 04	12.166
149	2 22 01	12.207	150	2 25 00	12.247
151	2 28 01	12.288	152	2 31 04	12.329
153	2 34 09	12.369	154	2 37 16	12.410
155	2 40 25	12.450	156	2 43 36	12.490
157	2 46 49	12.530	158	2 49 64	12.570
159	2 52 81	12.610	160	2 56 00	12.649
161	2 59 21	12.689	162	2 62 44	12.728
163	2 65 69	12.767	164	2 68 96	12.806
165	2 72 25	12.845	166	2 75 56	12.884
167	2 78 89	12.923	168	2 82 24	12.961
169	2 85 61	13.000	170	2 89 00	13.038
171	2 92 41	13.077	172	2 95 84	13.115
173	2 99 29	13.153	174	3 02 76	13.191
175	3 06 25	13.229	176	3 09 76	13.266
177	3 13 29	13.304	178	3 16 84	13.342
179	3 20 41	13.379	180	3 24 00	13.416
181	3 27 61	13.454	182	3 31 24	13.491
183	3 34 89	13.528	184	3 38 56	13.565
185	3 42 25	13.601	186	3 45 96	13.638
187	3 49 69	13.675	188	3 53 44	13.711
189	3 57 21	13.748	190	3 61 00	13.784
191	3 64 81	13.820	192	3 68 84	13.856
193	3 72 49	13.892	194	3 76 36	13.928
195	3 80 25	13.964	196	3 84 16	14.000
197	3 88 09	14.036	198	3 92 04	14.071
199	3 96 01	14.107	200	4 00 00	14.142

TABLE O (Contd.)

1	2	3	4	5	6
201	4 04 01	14.177	202	4 08 04	14.177
203	4 12 09	14.248	204	4 16 16	14.283
205	4 20 25	14.318	206	4 24 36	14.353
207	4 28 49	14.387	208	4 32 64	14.422
209	4 36 81	14.457	210	4 41 00	14.491
211	4 45 21	14.526	212	4 49 44	14.560
213	4 53 69	14.595	214	4 57 96	14.629
215	4 62 25	14.663	216	4 66 56	14.697
217	4 70 89	14.731	218	4 75 24	14.765
219	4 79 61	14.799	220	4 84 00	14.832
221	4 88 41	14.866	222	4 92 84	14.900
223	4 97 29	14.933	224	5 01 76	14.967
225	5 06 25	15.000	226	5 10 76	15.033
227	5 15 29	15.067	228	5 19 84	15.100
229	5 24 41	15.133	230	5 29 00	15.166
231	5 33 61	15.199	232	5 38 24	15.232
233	5 42 89	15.264	234	5 47 56	15.297
235	5 52 25	15.330	236	5 56 96	15.362
237	5 61 69	15.395	238	5 66 44	15.427
239	5 71 21	15.460	240	5 76 00	15.492
241	5 80 81	15.524	242	5 85 64	15.556
243	5 90 49	15.588	244	5 5 36	15.620
245	6 00 25	15.652	246	6 05 16	15.684
247	6 10 09	15.716	248	6 20 01	15.780
249	6 20 01	15.780	250	6 25 00	15.811
251	6 30 01	15.843	252	6 35 04	15.875
253	6 40 09	15.906	254	6 45 16	15.937
255	6 50 25	15.969	256	6 55 36	16.000
257	6 60 49	16.031	258	6 65 64	16.062
259	6 70 81	16.093	260	6 76 00	16.125
261	6 81 21	16.155	262	6 86 44	16.186
263	6 91 69	16.217	264	6 96 96	16.248
265	7 02 25	16.279	266	7 07 56	16.310
267	7 12 89	16.340	268	7 18 24	16.371
269	7 23 61	16.401	270	7 29 00	16.432

1	2	3	4	5	6
271	7 34 41	16.462	272	7 39 84	16.492
273	7 45 29	16.523	274	7 50 76	16.553
275	7 56 25	16.583	276	7 61 76	16.613
277	7 67 29	16.643	278	7 72 84	16.673
279	7 78 41	16.703	280	7 84 00	16.733
281	7 89 61	16.763	282	7 95 24	16.793
283	8 00 89	16.823	284	8 06 56	16.852
285	8 12 25	16.882	286	8 17 96	16.912
287	8 23 69	16.941	288	8 29 44	16.971
289	8 35 21	17.000	290	8 41 00	17.029
291	8 46 81	17.059	292	8 52 64	17.088
293	8 58 49	17.117	294	8 64 36	17.146
295	8 70 25	17.176	296	8 76 16	17.205
297	8 82 09	17.234	298	8 88 04	17.263
299	8 94 01	17.292	300	9 06 00	17.321
301	9 06 01	17.349	302	9 12 04	17.378
303	9 18 09	17.407	304	9 24 16	17.436
305	9 30 25	17.464	306	9 36 36	17.493
307	9 42 49	17.521	308	9 48 64	17.550
309	9 54 81	17.578	310	9 61 00	17.607
311	9 67 21	17.635	312	9 73 44	17.664
313	9 79 69	17.692	314	9 85 96	17.720
315	9 92 25	17.748	316	9 98 56	17.776
317	10 04 89	17.804	318	10 11 24	17.833
319	10 17 61	17.861	320	10 24 00	17.889
321	10 30 41	17.916	322	10 36 84	17.944
323	10 43 29	17.972	324	10 49 76	18.000
325	10 56 25	18.028	326	10 62 76	18.055
327	10 69 29	18.083	328	10 75 84	18.111
329	10 82 41	18.138	330	10 89 00	18.166
331	10 95 61	18.193	332	11 02 24	18.221
333	11 08 89	18.248	334	11 15 56	18.276
335	11 22 25	18.303	336	11 28 96	18.330
337	11 35 69	18.358	338	11 42 44	18.385
339	11 49 21	18.412	340	11 56 00	18.439
341	11 62 81	18.466	342	11 69 64	18.493

TABLE O (Contd.)

1	2	3	4	5	6
343	11 76 49	18.520	344	11 83 36	18.547
345	11 90 25	18.574	346	11 97 16	18.601
347	12 04 09	18.628	348	12 11 04	18.655
349	12 18 01	18.682	350	12 25 00	18.708
351	12 32 01	18.735	352	12 39 04	18.762
353	12 46 09	18.788	354	12 53 16	18.815
355	12 60 25	18.844	356	12 67 36	18.868
357	12 74 49	18.894	358	12 81 64	18.921
359	12 88 81	18.947	360	12 96 00	18.974
361	13 03 21	19.000	362	13 10 44	19.026
363	13 17 69	19.053	364	13 24 96	19.079
365	13 32 25	19.105	366	13 39 56	19.131
367	13 46 89	19.157	368	13 54 24	19.183
369	13 61 61	19.209	370	13 69 00	19.235
371	13 76 41	19.261	372	13 83 84	19.287
373	13 91 29	19.313	374	13 98 76	19.339
375	14 06 25	19.363	376	14 13 76	19.391
377	14 21 29	19.416	378	14 28 84	19.442
379	14 36 41	19.468	380	14 44 00	19.494
381	14 51 61	19.519	382	14 59 24	19.545
383	14 66 89	19.570	384	14 74 56	19.596
385	14 82 25	19.621	386	14 89 96	19.647
387	14 97 69	19.672	388	15 05 44	19.698
389	15 13 21	19.723	390	15 21 00	19.748
391	15 28 81	19.774	392	15 36 64	19.799
393	15 44 49	19.824	394	15 52 36	19.849
395	15 60 25	19.875	396	15 68 16	19.900
397	15 76 09	19.935	398	15 84 04	19.950
399	15 92 01	19.975	400	16 00 00	20.000
401	16 08 01	20.025	402	16 16 04	20.050
403	16 24 09	20.075	404	16 32 16	20.100
405	16 40 25	20.125	406	16 48 36	20.149
407	16 56 49	20.174	408	16 64 64	20.199
409	16 72 81	20.224	410	16 81 00	20.248

1	2	3	4	5	6
411	16 89 21	20.273	412	16 97 44	20.298
413	17 05 69	20.322	414	17 13 96	20.347
415	17 22 25	20.372	416	17 30 56	20.396
417	17 38 89	20.421	418	17 47 24	20.445
419	17 55 61	20.469	420	17 64 00	20.494
421	17 72 41	20.518	422	17 80 84	20.543
423	17 89 29	20.567	424	17 97 76	20.591
425	18 06 25	20.616	426	18 14 76	20.640
427	18 23 29	20.664	428	18 31 84	20.688
429	18 40 41	20.712	430	18 49 00	20.736
431	18 57 61	20.761	432	18 66 24	20.785
433	18 74 89	20.809	434	18 83 56	20.833
435	18 92 25	20.857	436	19 00 96	20.881
437	19 09 69	20.905	438	19 18 44	20.928
439	19 27 21	20.952	440	19 36 00	20.976
441	19 44 81	21.000	442	19 53 64	21.024
443	19 62 49	21.048	444	19 71 36	21.071
445	19 80 25	21.095	446	19 89 16	21.119
447	19 98 09	21.142	448	20 07 04	21.166
449	20 16 01	21.190	450	20 25 00	21.213
451	20 34 01	21.237	452	20 43 04	21.260
453	20 52 09	21.284	454	20 61 16	21.307
455	20 70 25	21.831	456	20 79 36	21.354
457	20 88 49	21.378	458	20 97 64	21.401
459	21 06 81	21.424	460	21 16 00	21.448
461	21 25 21	21.471	462	21 34 44	21.494
463	21 43 69	21.517	464	21 52 96	21.541
465	21 62 25	21.564	466	21 71 56	21.587
467	21 80 89	21.610	468	21 90 24	21.633
469	21 99 61	21.656	470	22 09 00	21.679
471	22 18 41	21.703	472	22 27 84	21.726
473	22 37 29	21.749	474	22 46 76	21.772
475	22 56 25	21.794	476	22 65 76	21.817
477	22 75 29	21.840	478	22 84 84	21.863
479	22 94 41	21.886	480	23 04 00	21.909
481	23 13 61	21.932	482	23 23 24	21.954
483	23 32 89	21.977	484	23 42 56	22.000

TABLE O (Contd.)

1	2	3	4	5	6
485	23 52 25	22 023	486	23 61 96	22 045
487	23 71 69	22 068	488	23 81 44	22 091
489	23 91 21	22 113	490	24 01 00	22 136
491	24 10 81	22 159	492	24 20 64	22 181
493	24 30 49	22 204	494	24 40 36	22 226
495	24 50 25	22 249	496	24 60 16	22 271
497	24 70 09	22 293	498	24 80 04	22 316
499	24 90 01	22 338	500	25 00 00	22 361
501	25 10 01	22 383	502	25 20 04	22 405
503	25 30 09	22 428	504	25 50 16	22 450
505	25 50 25	22 472	506	25 60 30	22 494
507	25 70 49	22 517	508	25 80 64	22 539
509	25 90 81	22 561	510	26 01 00	22 583
511	26 11 21	22 605	512	26 21 44	22 627
513	26 31 69	22 650	514	26 41 96	22 672
515	26 52 25	22 694	516	26 62 56	22 716
517	26 72 89	22 738	518	26 83 24	22 760
519	26 93 61	22 782	520	27 04 00	22 804
521	27 14 41	22 825	522	27 24 84	22 847
523	27 35 29	22 869	524	27 45 75	22 891
525	27 56 25	22 913	526	27 66 76	22 935
527	27 77 29	22 956	528	27 87 84	22 978
529	27 98 41	23 000	530	28 09 00	23 022
531	28 19 61	23 043	532	28 30 24	23 065
533	28 40 89	23 087	534	28 51 56	23 108
535	28 62 25	23 130	536	28 72 96	23 152
537	28 83 69	23 173	538	28 94 44	23 195
539	29 05 21	23 216	540	29 16 00	23 238
541	29 26 81	23 259	542	29 37 64	23 281
543	29 48 49	23 302	544	29 59 36	23 324
545	29 70 25	23 345	546	30 01 00	23 367
547	29 92 09	23 388	548	30 23 04	23 409
549	30 14 01	23 431	550	30 45 00	23 452
551	30 36 01	23 473	552	30 47 04	23 495

1	2	3	4	5	6
553	30 58 09	23.516	554	30 59 16	23.537
555	30 80 25	23.558	556	30 91 36	23.580
557	31 02 49	23.601	558	31 13 64	23.622
559	31 24 81	23.643	560	31 36 00	23.664
561	31 47 21	23.685	562	31 58 44	23.707
563	31 69 69	23.728	564	31 80 96	23.749
565	31 92 25	23.770	566	32 03 56	23.791
567	32 14 89	23.812	568	32 26 24	23.833
569	32 37 61	23.854	570	32 49 00	23.875
571	32 60 41	23.896	572	32 71 84	23.917
573	32 83 29	23.937	574	32 94 76	23.958
575	33 06 25	23.979	576	33 17 76	24.000
577	33 29 29	24.021	578	33 40 84	24.042
579	33 52 41	24.062	580	33 64 00	24.083
581	33 75 61	24.104	582	33 87 24	24.125
583	33 98 89	24.145	584	34 10 56	24.166
585	34 22 25	24.187	586	34 33 96	24.207
587	34 45 69	24.228	588	34 57 44	24.249
589	34 69 21	24.269	590	34 81 00	24.290
591	34 92 81	24.310	592	35 04 64	24.331
593	35 16 49	24.352	594	35 28 36	24.372
595	35 40 25	24.393	596	35 52 16	24.413
597	35 64 09	24.434	598	35 76 04	24.454
599	35 88 01	24.474	600	36 00 00	24.495
601	36 12 01	24.515	602	36 24 04	24.536
603	36 36 01	24.556	604	36 48 16	24.576
605	36 60 25	24.597	606	36 72 36	24.617
607	36 84 49	24.637	608	36 96 64	24.658
609	37 08 81	24.678	610	37 21 00	24.698
611	37 33 21	24.718	612	37 45 44	24.739
613	37 57 69	24.759	614	37 69 96	24.779
615	37 82 25	24.799	616	37 94 56	24.819
617	38 06 89	24.839	618	38 19 24	24.860
619	38 31 61	24.880	620	38 44 00	24.900
621	38 56 41	24.920	622	38 68 84	24.940
623	38 81 29	24.960	624	38 93 76	24.980

TABLE O (Contd.)

1	2	3	4	5	6
625	39 06 25	25.000	626	39 18 76	25.020
627	39 31 29	25.040	628	39 43 84	25.060
629	39 56 41	25.080	630	39 69 00	25.100
631	39 81 61	25.120	632	39 94 24	25.140
633	40 06 89	25.159	634	40 19 56	25.179
635	40 32 25	25.199	636	40 44 96	25.219
637	40 57 69	25.239	638	40 70 44	25.259
639	40 83 21	25.278	640	40 96 00	25.298
641	41 08 81	25.318	642	41 21 64	25.338
643	41 34 49	25.357	644	41 47 36	25.377
645	41 60 25	25.397	646	41 73 16	25.417
647	41 86 09	25.436	648	41 99 04	25.456
649	42 12 01	25.475	650	42 25 00	25.495
651	42 38 01	25.515	652	42 51 04	25.534
653	42 64 09	25.554	654	42 77 16	25.573
655	42 90 25	25.593	656	43 03 36	25.612
657	43 16 49	25.632	658	43 29 64	25.652
659	43 42 81	25.671	660	43 56 00	25.690
661	43 69 21	25.710	662	43 82 44	25.729
663	43 95 69	25.749	664	44 08 96	25.768
665	44 22 25	25.788	666	44 35 56	25.807
667	44 48 89	25.826	668	44 62 24	25.846
669	44 75 61	25.865	670	44 89 00	25.884
671	45 02 41	25.904	672	45 15 84	25.923
673	45 29 29	25.942	674	45 42 76	25.962
675	45 56 25	25.981	676	45 69 76	26.000
677	45 83 29	26.019	678	45 96 84	26.038
679	46 10 41	26.058	680	46 24 00	26.077
681	46 37 61	26.096	682	46 51 24	26.115
683	46 64 89	26.134	684	46 78 56	26.153
685	46 92 25	26.173	686	47 05 96	26.192
687	47 19 69	26.211	688	47 33 44	26.230
689	47 47 21	26.249	690	47 61 00	26.268
691	47 74 81	26.287	692	47 88 64	26.306
693	48 02 49	26.325	694	48 16 36	26.344

1	2	3	4	5	6
695	48 30 25	26 363	696	48 41 16	26 382
697	48 58 09	26 401	698	48 72 04	26 420
699	48 86 01	26 439	700	49 00 00	26 458
701	49 14 01	26 476	702	49 28 04	26 495
703	49 42 09	26 514	704	49 56 16	26 533
705	49 70 25	26 552	706	49 84 36	26 571
707	49 98 49	26 589	708	50 12 64	26 608
709	50 26 81	26 627	710	50 41 00	26 646
711	50 55 21	26 665	712	50 69 44	26 683
713	50 83 69	26 702	714	50 97 96	26 721
715	51 12 25	26 739	716	51 26 56	26 758
717	51 40 89	26 777	718	51 55 24	26 796
719	51 69 61	26 814	720	51 84 00	26 833
721	51 98 41	26 851	722	52 12 84	26 870
723	52 27 29	26 889	724	52 41 76	26 907
725	52 56 25	26 926	726	52 70 76	26 944
727	52 85 29	26 963	728	52 99 84	26 981
729	53 14 41	27 000	730	53 29 00	27 019
731	53 43 61	27 037	732	53 58 24	27 055
733	53 72 89	27 074	734	53 87 56	27 092
735	54 02 25	27 111	736	54 10 96	27 129
737	54 31 69	27 148	738	54 46 44	27 166
739	54 61 21	27 185	740	54 76 00	27 208
741	54 90 81	27 221	742	55 05 64	27 240
743	55 20 49	27 258	744	55 35 36	27 276
745	55 50 25	27 295	746	55 65 16	27 313
747	55 80 09	27 331	748	55 95 04	27 350
749	56 10 01	27 368	750	56 25 00	27 386
751	56 40 01	27 404	752	56 55 04	27 423
753	56 70 09	27 441	754	56 85 16	27 459
755	57 00 25	27 477	756	57 15 36	27 495
757	57 30 49	27 514	758	57 45 64	27 532
759	57 60 81	27 550	760	57 76 00	27 568
761	57 91 21	27 586	762	58 06 44	27 604
763	58 21 69	27 622	764	58 36 96	27 641
765	58 52 25	27 659	766	58 67 56	27 677

TABLE O (Contd.)

1	2	3	4	5	6
767	58 82 89	27.695	768	58 98 24	27.713
769	59 13 61	27.731	770	59 29 00	27.749
771	59 44 41	27.767	772	59 59 84	27.785
773	59 75 29	27.803	774	59 90 76	27.821
775	60 06 25	27.839	776	60 21 76	27.857
777	60 37 29	27.875	778	60 52 84	27.893
779	60 68 41	27.911	780	60 84 00	27.928
781	60 99 61	27.946	782	61 15 24	27.964
783	61 30 89	27.982	784	61 46 56	28.000
785	61 62 25	28.018	786	61 77 96	28.036
787	61 93 69	28.054	788	62 09 44	28.071
789	62 25 21	28.089	790	62 41 00	28.107
791	62 56 81	28.125	792	62 72 64	28.142
793	62 88 49	28.160	794	63 04 36	28.178
795	63 20 25	28.196	796	63 36 16	28.213
797	63 52 09	28.231	798	63 68 04	28.249
799	63 84 01	28.267	800	64 00 00	28.284
801	64 16 01	28.302	802	64 32 04	28.320
803	64 48 09	28.337	804	64 64 16	28.355
805	64 80 25	28.373	806	64 96 36	28.390
807	65 12 49	28.408	808	65 28 64	28.425
809	65 44 81	28.443	810	65 61 00	28.460
811	65 77 21	28.478	812	65 93 44	28.496
813	66 09 69	28.513	814	66 25 96	28.531
815	66 42 25	28.548	816	66 58 56	28.566
817	66 74 89	28.583	818	66 91 24	28.601
819	67 07 61	28.618	820	67 24 00	28.636
821	67 40 41	28.653	822	67 56 54	28.671
823	67 73 29	28.688	824	67 89 76	28.705
825	68 06 25	28.723	826	68 22 76	28.740
827	68 39 29	28.758	828	68 55 84	28.775
829	68 72 41	28.792	830	68 89 00	28.810
831	69 05 61	28.827	832	69 22 24	28.844
833	69 38 89	28.862	834	69 55 56	28.879
835	69 72 25	28.896	836	69 88 96	28.914

1	2	3	4	5	9
837	70 05 69	28.931	838	70 22 44	28.948
839	70 39 21	28.965	840	70 56 00	28.983
841	70 72 81	29.000	842	70 89 64	29.017
843	71 06 49	29.034	844	71 23 36	29.052
845	71 40 25	29.069	846	71 57 16	29.086
847	71 74 09	29.103	848	71 91 04	29.120
849	72 08 01	29.138	850	72 25 00	29.155
851	72 42 01	29.172	852	72 59 04	29.189
853	72 76 09	29.206	854	72 93 16	29.223
855	73 10 25	29.240	856	73 27 36	29.257
857	73 44 49	29.275	858	73 61 64	29.292
859	73 78 81	29.309	860	73 96 00	29.326
861	74 13 21	29.343	862	74 30 44	29.360
863	74 47 69	29.377	864	74 64 96	29.394
865	74 82 25	29.411	866	74 99 56	29.428
867	75 16 89	29.445	868	75 34 24	29.462
869	75 51 61	29.479	870	75 69 00	29.496
871	75 86 41	29.513	872	76 03 84	29.530
873	76 21 29	29.547	874	76 38 76	29.563
875	76 56 25	29.580	876	76 73 76	29.597
877	76 91 29	29.614	878	77 08 84	29.631
879	77 26 41	29.648	880	77 44 00	29.665
881	77 61 61	29.682	882	77 79 24	29.698
883	77 96 89	29.715	884	78 14 56	29.732
885	78 32 25	29.749	886	78 49 96	29.766
887	78 67 69	29.783	888	78 85 44	29.799
889	79 03 21	29.816	890	79 21 00	29.833
891	79 38 81	29.850	892	79 56 64	29.866
893	79 74 49	29.883	894	79 92 36	29.900
895	80 10 25	29.916	896	80 28 16	29.933
897	80 46 09	29.950	898	80 64 04	29.967
899	80 82 01	29.983	900	81 00 00	30.000
901	81 18 01	30.017	902	81 36 04	30.033
903	81 54 09	30.050	904	81 72 16	30.067
905	81 90 25	30.083	906	82 08 36	30.100
907	82 26 49	30.116	908	82 44 64	30.133

TABLE O (Contd.)

1	2	3	4	5	6
909	82 62 81	30.150	910	82 81 00	30.166
911	82 99 21	30.183	912	83 17 44	30.199
913	83 35 69	30.216	914	83 53 96	30.232
915	83 72 25	30.249	916	83 90 56	30.265
917	84 08 89	30.282	918	84 27 24	30.299
919	84 45 61	30.315	920	84 64 00	30.332
921	84 82 41	30.348	922	85 00 84	30.364
923	85 19 29	30.381	924	85 37 76	30.397
925	85 56 25	30.414	926	85 74 76	30.430
927	85 93 29	30.447	928	86 11 84	30.463
929	86 30 41	30.480	930	86 49 00	30.496
931	86 67 61	30.512	932	86 86 24	30.529
933	87 04 89	30.545	934	87 23 56	33.561
935	87 42 25	30.578	936	87 60 96	30.594
937	87 79 69	30.610	938	87 98 44	30.627
939	88 17 21	30.643	940	88 36 00	30.659
941	88 54 81	30.676	942	88 73 64	30.692
943	88 92 49	30.708	944	89 11 36	30.725
945	89 30 25	30.741	946	89 49 16	30.757
947	89 68 09	30.773	948	89 87 04	30.790
949	90 06 01	30.806	950	90 25 00	30.822
951	90 44 01	30.838	952	90 63 04	30.854
953	90 82 09	30.871	954	91 01 16	30.887
955	91 20 25	30.903	956	91 39 36	30.919
957	91 58 49	30.935	958	91 77 64	30.952
959	91 96 81	30.968	960	92 16 00	30.984
961	92 35 21	31.000	962	92 54 44	31.016
963	92 73 69	31.032	964	92 92 96	31.048
965	93 12 25	31.064	966	93 31 50	31.081
967	93 50 89	31.097	968	93 70 24	31.113
969	93 89 61	31.129	970	94 09 00	31.145
971	94 28 41	31.161	972	94 47 84	31.177
973	94 67 29	31.193	974	94 86 76	31.209
975	95 06 25	31.225	976	95 25 76	31.241

1	2	3	4	5	6
977	95 45 29	31.257	978	95 64 84	31.273
979	95 84 41	31.289	980	96 04 00	31.305
981	96 23 61	31.321	982	96 43 24	31.337
983	96 62 89	31.353	984	96 82 56	31.369
985	97 02 25	31.385	986	97 21 96	31.401
987	97 41 69	31.417	988	97 51 44	31.432
989	97 81 21	31.448	990	98 01 00	31.464
991	98 20 81	31.480	992	98 40 64	31.496
993	98 60 49	31.512	994	98 80 36	31.528
995	99 00 25	31.544	996	99 20 16	31.559
997	99 40 09	31.575	998	99 60 04	31.591
999	99 80 01	31.607	1000	100 00 00	31.623

ANSWERS TO EXERCISES FOR PRACTICE

CHAPTER 1

- 1.4 (a) Ratio (b) Ratio (c) Interval (d) Nominal
(e) Nominal (f) Ratio (g) Ordinal (h) Ratio
(i) Interval.

CHAPTER 3

- 3.1 (a) $M=5$; $Mdn=5$; $Mo=2$
(b) $M=14.86$; $Mdn=14.00$; $Mo=14.00$
(c) $M=13.17$; $Mdn=13.50$; $Mo=15.00$
- 3.2 (a) $M=48.80$; $Mdn=49.50$; $Mo=50.90$
(b) $M=151.70$; $Mdn=154.76$; $Mo=160.88$
(c) $M=27.80$; $Mdn=27.58$; $Mo=27.14$
(d) as in a, b, c above.

CHAPTER 4

- 4.1 $Var=458.72$; $SD=21.42$;
4.3 $AD=15.92$; $Q=11.305$

CHAPTER 5

- 5.5 *T Scores*, 74.8; 69.4; 65.6; 61.8; 58.1;
53.2; 45.9; 37.1, 27.8
(For top nine intervals)

5.6 *Stanine* *Exact Limits of Intervals*

9	44.43—
8	37.53—44.43
7	31.08—37.53
6	26.44—31.08
5	22.85—26.44
4	20.36—22.85
3	17.39—20.36
2	14.97—17.39
1	— 14.97

5.7 39.5, 49.5, 59.5, 19.5, 69.5, 79.5

5.9

<i>Group</i>	<i>z Scores</i>	
	History	Maths
I	+ .25	— .2
II	+ 1.63	+ .3
III	— 1.0	— .7

CHAPTER 6

6.1 (a) $1/4$ (b) $1/4$ (c) $3/4$ (d) $3/4$ (e) $1/2$ 6.2 (a) $1/2$ (b) $7/8$ (c) $1/8$ (d) $1/8$ (e) $3/8$ 6.3 $5/6$ 6.4 (a) $2^6=64$ (b) $(\frac{1}{2}+\frac{1}{2})^6$ (c) $R^6+6R^5W+15R^4W^2+20R^3W^3+15R^2W^4+6RW^5+W^6$

Check : The total of numerical coefficients.

$$1+6+15+20+15+6+1=64$$

(d) (i) $22/64$ (ii) $1/64$ (iii) $20/64$ (e) 22; 1; 20.

6.5 (a) 16; (b) 16; (c) 60; (d) 4.

6.6 $\frac{5!}{3! 2!} (.4)^3 (.6)^2 = .2304$.

6.8 (a) 118 (b) 45.36; 40.00 (c) 11, 68, 171, 171, 68, 11

6.10 .26; 1.28

CHAPTER 7

- 7.4 $r = .87$
7.5 $\rho = .76$
7.6 $r = -.50$
7.7 (a) No difference (b) No change
7.8 $r_{bis} = .19$
7.9 $r_{pbis} = .84$
7.10 $r_t = .139$
7.11 Phi Coefficient = .20

CHAPTER 8

- 8.1 $\sigma_M = .25$; Confidence Intervals .99 : 24.955—26.245
.95 : 25.11 —26.09
- 8.2 $\sigma_M = 2.04$; Confidence Intervals .99 : 129.31—140.69
.95 : 130.80—139.20
- 8.3 (a) 74 in 100; (b) 32 in 100; (c) less than 1 in 100
- 8.4 $\sigma_p = 3.96$; .99 limits : 44.78—65.22
- 8.5 $\sigma_{M, \text{in}} = .52$; Confidence Intervals: .99 : 22.95—26.04
.95 : 23.48—25.52
- 8.6 $\sigma_\sigma = .51$; Confidence Intervals : .99 : 8.88—11.52
.95 : 9.20—11.20
- 8.7 $\sigma_z = .17$; Confidence Intervals : .99 : .55—.855
.95 : .47—.88

CHAPTER 9

- 9.1 $t=4.322$, sig.
9.2 $t=6.60$; sig.
9.3 $t=10.32$; sig.
9.5 (a) $\sigma_{SE}=.034$; $t=2.14$: sig at .05 level
(b) $\sigma_{SE}=.029$; $t=5.52$; sig. at .01 level
9.6 $SE=1.14$; $t=7.98$; sig. at .01 level

9.7 $SE=1.046$; $t=4.78$; sig. at .01 level

9.8 (a) .69; $-.62$; .26; -1.26 .

(c) R.A. Fisher.

9.9 (a) $SE=.156$; $t=1.73$; n.s.

(b) $SE=.158$; $t=7.00$; sig. at .01

(c) $SE=.127$; $t=5.8$; sig. at .01

(d) $SE=.07$; $t=5.14$; sig. at .01

(e) $SE=.141$; $t=5.17$; sig. at .01

9.10 $SE=.141$; $t=3.90$; sig. at .01

CHAPTER 10

10.1 $\chi^2=16.17$; sig. at .01 level

10.2 $\chi^2=31.12$; sig. at .01 level

10.3 $\chi^2=4.20$; not significant

10.4 $\chi^2=.41$; H_0 accepted.

10.5 $\chi^2=8.1$; sig. .01 level

10.6 $z = .6057$

10.7 $D=.592$ (Critical D at .05 = .328; at .01 = .396)

$\chi^2=24.034$

CHAPTER 11

11.4 (a)	Source	df	SS	MS	F	
	Between	3	46.94	15.65	8.46	sig. at 01
	Within	13	24.00	1.846		
(b)	Source	df	SS	MS	F	
	Between	4	110.0	27.4	4.365	sig. at .05
	Within	20	126.0	6.3		

(c)	Source	df	SS	MS	F
	Drug	2	29.40	14.70	10.13 sig at .01
	Sex	1	4.03	4.03	2.78 n.s.
	Interaction	2	8.07	4.03	2.78 n.s.
	Within	24	34.80	1.45	

(d)	Source	df	SS	MS	F
	Method	2	1.57	.79	<1.0 n.s.
	Teacher	2	0.00	0.00	0.00 n.s.
	Interaction	4	32.43	8.11	2.01 n.s.
	Within	18	72.67	4.04	

(e)	Source	df	SS	MS	F
	Condition	3	90.5	30.16	4.65 sig. at .05
	Sex	1	72.0	72	11.7 sig. at .01
	Interaction	3	7.0	2.33	.35 n.s.
	Within	24	156.0	6.5	

11.5 Significant Pairs of Means

$M_1 - M_2$ sig. at .01

$M_1 - M_3$ sig. at .05

$M_2 - M_4$ sig. at .05

$M_2 - M_5$ sig. at .01

$M_3 - M_5$ sig. at .05

11.6 (a) Between: 3; Within: 56; Total: 59

(b) Between: 2; Within: 27; Total: 29

(c) A : 2; B : 1; C : 3; A \times B; 2; A \times C : 8; B \times C : 3;
A \times B \times C : 6; Within : 94; Total : 119.

11.9 (A)

	M_1	M_2		M_1	M_2	
S_1	17	9	13	S_1	13	13
S_2	17	9	13	S_2	13	13
	17	9	13		13	13

12.2 Correlation within.

12.3 Adjusted Y mean will be smaller.

CHAPTER 13

13.4 (i) .67 (ii) .82

13.5 .81

13.7 Difficulty Index : .70, .43, .38, .45

Discrimination Index: .20, .05, .24, .58.

CHAPTER 14

14.1 $Y' = X + 2$

$$X' = .76Y - 0.56$$

14.2 (a) $Y' = .84X + 0.8$

$$X' = .583Y + 65.87$$

(b) 135.34; 118.34; 138.745; and 141.66.

(c) 103.28; 97.4; 118.4; 114.2; 111.68; and 108.32.

14.3 $Y' = .9X + 1.2$

14.5 (a) $X'_1 = .69X_3 + 1.072X_4 - 24.27$

(b) 76.46; 76.69; 78.37.

BIBLIOGRAPHY

- Bancroft, T.A., 1968: *Topics in Intermediate Statistical Methods*, The Iowa State University Press, Ames; Iowa.
- Bradley, James V., 1968: *Distribution-free Statistical Tests*, Prentice-Hall, Inc., Englewood Cliffs, N.J.
- Cochran, William G., and Gertrude M. Cox, 1957: *Experimental Designs*, 2d ed., John Wiley & Sons, Inc., New York.
- Cox, D.R., 1958: *Planning of Experiments*, John Wiley & Sons, Inc., New York.
- Cronbach, L.J., 1957: The Two Disciplines of Scientific Psychology, *The American Psychologist*, 12:671-684.
- Dayton, C. Mitchell, 1970: *Design of Educational Experiments*, McGraw-Hill Book Company, New York.
- Dubois Phillips H., 1970: Varieties of Psychological Test Homogeneity, *The American Psychologist*, 25:532-536.
- Duncan, D.B., 1955: Multiple Range and Multiple F-tests, *Biometrics*, 11:1-42.
- Edwards, Allen L., 1967: *Statistical Methods*, 2d ed., Holt, Rinehart and Winston, Inc., New York.
- , 1968: *Experimental Design in Psychological Research*, 3d ed., Holt, Rinehart and Winston, Inc. New York.
- Ferguson, George A., 1941: *The Reliability of Mental Tests*, University of London Press, Ltd., London.
- , 1965: *Nonparametric Trend Analysis*, McGill University Press, Montreal.
- , 1976: *Statistical Analysis in Psychology and Education*, McGraw Hill, New York.
- Finney, D.J., 1948: The Fisher-Yates Test of Significance in 2×2 Contingency Tables, *Biometrika*, 35:145-156.

- Fisher, R.A., 1970: *Statistical Methods for Research Workers*, 14th ed. Hafner Publishing Company, Inc., New York,
- , and F. Yates, 1963: *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., Hafner Publishing Company, Inc., New York.
- Freund, John E., 1973: *Modern Elementary Statistics*, 4th ed., Prentice Hall Inc., Englewood Cliffs, N.J.
- Friedman, M., 1937: The Use of Banks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, *Journal of the American Statistical Association*, 32:675-701.
- Garrett, Henry E. 1973: *Statistics in Psychology and Education*, Vakils, Feffer and Simon, Bombay.
- Glass, Gene V., and Julian C. Stanley, 1970: *Statistical Methods in Education and Psychology*, Prentice-Hall, Inc, Englewood Cliffs, N.J.
- Guilford, J.P., and Benjamin Fruchter, 1973: *Fundamental Statistics in Psychology and Education*, 5th ed., McGraw-Hill Book Company New York.
- Gulliksen, H., 1950: *Theory of Mental Tests*, John Wiley & Sons, Inc., New York.
- Hays, W.L., 1973: *Statistics for the Social Sciences*, 2d ed., Holt, Rinehart and Winston, Inc., New York
- Keeping, E.S., 1962: *Introduction to Statistical Inference*, D Van Nostrand Company, Inc. Princeton, N.J.
- Kendall, M.G., 1970: *Rank Correlation Methods*, 4th ed., Charles Griffin & Company, Ltd., London.
- , and Alan Stuart, 1968-1973: *The Advanced Theory of Statistics*, 3 Vols, Hafner Publishing Company, Inc., New York.
- Kenney, John F., and E.S. Keeping, 1958: *Mathematics of Statistics*, part 1, 3d ed., D. Van Nostrand Company, Inc., Princeton, N.J.
- Kerlinger, Fred N., and Elazar J. Pedhazur, 1973: *Multiple Regression in Behavioral Research*, Holt, Rinehart and Winston, Inc., New York.

- Kruskal, W.H., and W.A. Wallis. 1952: Use of Ranks in one-criterion Variance Analysis, *Journal of the American Statistical Association*, 47:583-621.
- Lord, Frederic M., 1955a: Estimating Test Reliability, *Educational and Psychological Measurement*, 15:325-336.
- , 1955b: Sampling Fluctuations Resulting from the Sampling of Test Items, *Psychometrika*, 20:1-22.
- , and Melvin R. Novick, 1968: *Statistical Theories of Mental Test Scores*, Addison Wesley Publishing Company, Inc., Reading, Mass.
- McNemar, Quinn, 1947: Note on the Sampling Error of the Differences between Correlated Proportions or Percentages, *Psychometrika*, 12:153-157.
- Ray, William S., 1960: *An Introduction to Experimental Design*, The Macmillan Company, New York.
- Scheffe, H., 1953: A Method for Judging All Contrasts in the Analysis of Variance, *Biometrika*, 40:87-104.
- , 1959: *The Analysis of Variance*, John Wiley & Sons, Inc., New York.
- Siegel, Sidney 1956: *Nonparametric Statistics*, McGraw-Hill Book Company, New York.
- Snedecor, George W., and William G. Cochran, 1967: *Statistical Methods*, 5th ed., The Iowa State University Press Ames, Iowa.
- Tukey, John W., 1949: Comparing Individual Means in the Analysis of Variance, *Biometrics*, 5:99-114.
- Walker, Helen M., and Joseph Lev, 1953: *Statistical Inference*, Holt, Rinehart and Winston, Inc., New York.
- Winer, B.J., 1971: *Statistical Principles in Experimental Design*, 2d ed., McGraw-Hill Book Company, New York.

INDEX

- a Coefficient, 300-18
- Achievement Test, 77
- Age Norms, 79
 - Def., examples, merits, 79
- Alternative Hypothesis, 184
- Analysis of Covariance 270-83
 - Assumptions, 281
 - Adjusted means, 280
 - Computation, 272-76
 - General Uses, 281-82
 - Notation, 276-81
- Analysis of Variance, 239-69
 - Assumptions, 264-65
 - Deviation score method, 243-48
 - General uses, 265-66
 - Interaction, 239, 255, 258, 260-64
 - Limitations, 265-66
 - One way or single classification, 241-51
 - Post ANOVA t test, 250-51
 - Rationale, 239-40
 - Raw score method, 248-50, 253-56
 - Relationship with t, 266
- Anthropometrical Data, 130
- Arithmetic Ability, 79
- Arithmetic Test, 77
- Assumed Mean, AM, 68
- Assumptions
 - of ANOVA, 264-65
 - of ANCOVA, 281
 - of chi square, 215
 - of rho, 147
- b Coefficient, 300-18
- Binomial Distribution, 103, 108-113
 - Binomial coefficients, 111
 - Mean of, 113
 - Pascal's Triangle, 111, 124
 - SD of, 113
- Biserial Correlation, 155-57
- Bivariate Distribution; 300
- Causation, 153
- Class Intervals, 10-36
 - Assumptions of, 15
 - Exact limits of, 17
 - Mid points of, 17
 - Number of, 15
- Chi Square χ^2 , 203-17
 - Additivity, 216-17
 - Assumptions, 215
 - Degrees of freedom, 205-6
 - Distribution, 204
 - Equal probability, 206-08
 - Independence test, 208-09
 - Normality test, 209-11
 - Percentages, 214-15
 - Yates' correction, 213
 - 2x2 Contingency tables, 211-13
- Coefficient
 - of Equivalence, 287
 - Reliability, 284
 - of Stability, 285
- Combined Mean, 44-45
- Confidence Intervals, 169-71

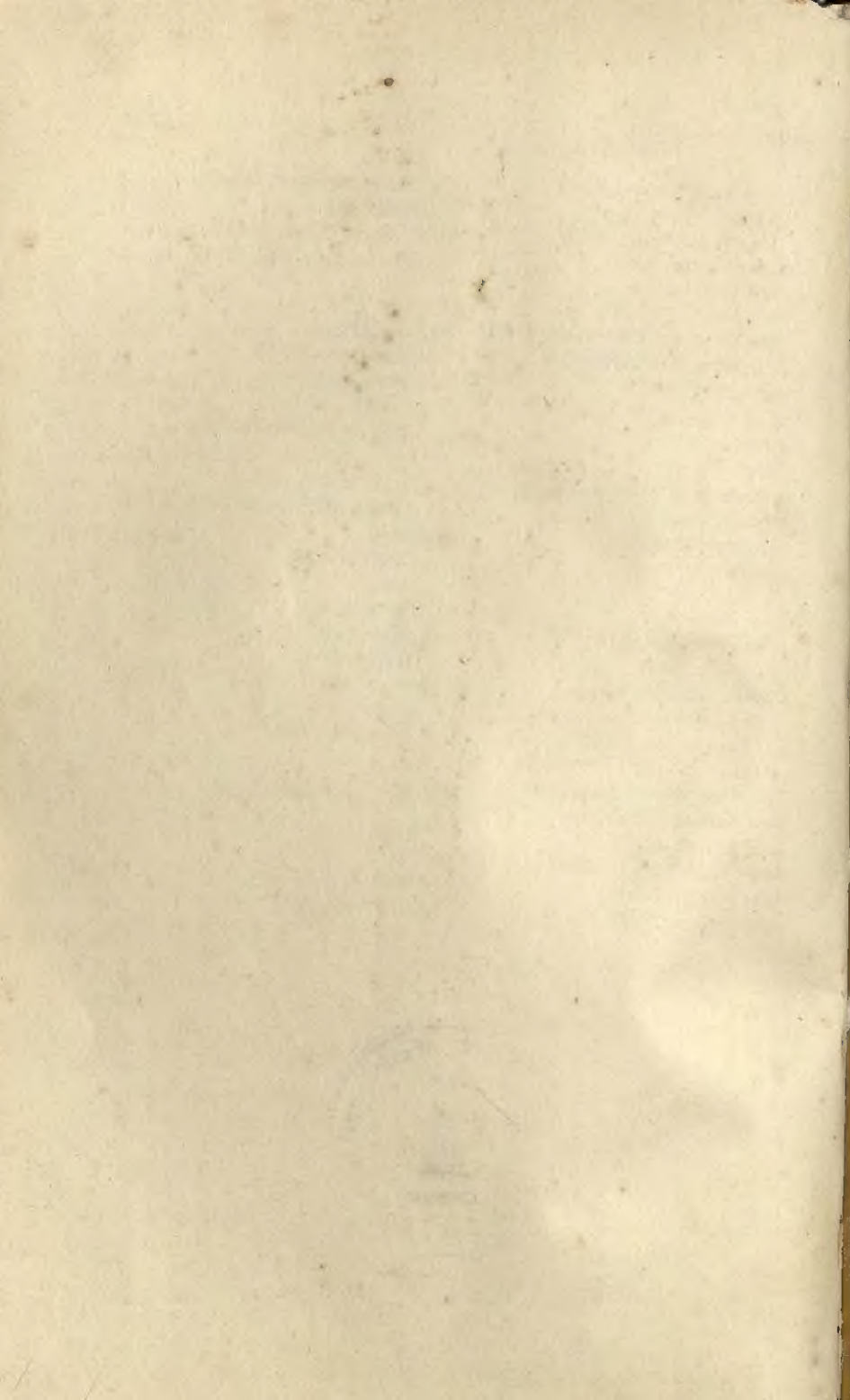
- Correlational Techniques, 142-65
 - Biserial correlation, 155-57
 - Canonical, 317
 - Coefficient of alienation, 151
 - Coefficient of determination, 151
 - Maximal, 143
 - Multiple, 312-14
 - Perfect, 143
 - Phi coefficient, 162-63
 - Point biserial correlation, 157-59
 - Product Moment Correlation, 143-47
 - assumptions, 154; deviation score, 144
 - difference formula, 146; effect of origin, 152;
 - effect of unit 152; raw score method, 144;
 - verbal interpretation, 155
 - Scatter plots, 142
 - Rank order correlation, 147-51
 - Tetrachoric correlation, 159-61
- Cumulative Frequency Curve, 29-31
- Cumulative Percentage Curve, 31-32
- Curvilinearity of Regression, 154
- Darwin's Theory of Evolution, 300
- Degrees of Freedom, 172-73
- Difference Method, 190
- Directional Test, 200
- Doolittle Solution, 317
- Equation of a Straight Line, 301-03
 - Slope a , 303
 - Y-intercept, 303
- Equivalent Groups, 189
- Errors
 - Chance, 130
 - of Observation, 130
 - of Sampling, 154
 - of Measurement, 154
- Estimates, 3
- F Distribution, 239-83
- F Formula, 240
- F Ratio, 239-83
- Fiduciary Limits, 169-71
- Fisher, R.A., 170, 240
- Fisher's Z. Function, 180-81
- Flanagan's Abac, 298
- Flanagan, J.C., 298
- Frequency, 10-23
 - Def., 10
 - Relative f , 12
 - Relative frequency distribution, 12
- Frequency Distributions, 10-33
 - Def., 10
 - Development, 11
 - Steps, 14-16
 - General Rules, 16-17
- Frequency Polygon, 24-27
 - Smoothed, 28
- Galton, Sir Francis, 300
- Grade Norms, 79-80
- Graphic Representation of Data, 20-34
- Histograms, 22-24
- Hypergeometric, 102
- Independent Means, 186
- Independent Samples, 187
- Indices of Discrimination, 298
- Inferential Statistics, 130
- Intelligence, 130
- Interaction, 139-55, 258, 260-64
 - disordinal, 262
 - ordinal, 262
- Item Analysis, 295-98
 - Correction for guessing, 296-97
 - Item difficulty, 295-96
 - Item discrimination, 297-98
- Karl Pearson
- K-S Test, 227-32
- Kuder Richardson Formula, 288
- Kurtosis, 115, 132-38
 - Formula, 133
 - Importance, 136
 - Leptokurtic, 132

- Meanings, 132
- Mesokurtic, 132
- Moments, 133-36
- Platykurtic, 132
- Significance, 136
- Mathematical Consequences, 130
- Mean, M , 34-45, 65-67, 115, 166-76, 183-202
 - Centre of gravity, 41
 - Combined mean, 44-45
 - Def., of, 35
 - Long method, 36-37
 - Properties of, 41-45
 - Significance of, 166-76
 - Significance of difference, 183-202
 - Short method, 38-40
- Measures of Central Tendency, 34-57
- Mean Deviation, 115
- Measures of Variability, 59-76
 - Average deviation (AD), 61-62
 - Range, 60-61
 - Semi-inter quartile range, Q , 70-72
 - Standard deviation, 61-75
 - Variance, 64-75
- Measures of Relative Standing, 77-102
- Measurement Scales, 4-8
 - Absolute zero point, 5
 - Equal intervals, 5
 - Interval scale, 6-7, 235
 - Magnitude, 5
 - Nominal, 5-7, 234
 - Ordinal, 6-7, 234
 - Ratio, 7-8
- Median (M_d), 45-52
 - Comparison with mean and mode, 54-57
 - Def. of, 45
 - Frequency distribution, 48-52
 - Guidelines for use, 56
 - Standard error of, 176-78
 - Steps, 49-50
 - Ungrouped data, 45-47
- Moments, 133-36
 - First, 134
 - Fourth, 134
 - Kurtosis, 133-36
 - Second, 134
 - Skewness, 133-36
 - Third, 134
- Multiple Correlation, R , 312-14
- Multiplication Theorem, 107
- Multivariate Factorial Designs, 70
- Mutually Exclusive Events, 105, 106
- Non-Critical Region, 200
- Non-Parametric Methods, 203-38
 - Kolmogorov Smirnov (K-S) test, 227-32
 - Median test, 223-25
 - Run test, 225-27
 - Sign test, 218-22
- Normal Curve, 96, 113-31
 - Areas, 118-19
 - Cases within score limits, 119
 - Comparison of distribution, 125
 - Division into sub-groups, 127-29
 - Equation, 115, 116
- Normal Curve
 - Importance, 130
 - Maximum ordinate, 115
 - Percentage above a score, 122
 - Percentage below a score, 123
 - Points of inflection, 115
 - Problems, 119-31
 - Properties, 114-15
 - Unit normal curve, 116
- Normalized Standard Scores, 96
- Norms, 77-81
 - Age norms, 78-79
 - Grade norms, 78-80
 - Percentile norms, 78-93
 - Standard scores, 78, 93-95
- Null Hypothesis, H_0 , 183-84
- Parameter, 3
- Partial Regression Coefficient, 314-16
- Pearson, Egon S., 136
- Pearson, Karl, 143

- Percentile, 80-93
 - from Grouped data, 86-89
 - from Ogive, 92-93
 - from Ungrouped data, 86-89
- Percentile Ranks
 - from Grouped data, 89-91
 - from Ogive, 91-93
 - from Ungrouped data, 84-86
- Permutations, 107-08
- Phi Coefficient, 162-63
- Point Biserial Correlation, 157-59
- Poisson, 102
- Principle of Least Squares, 42-43
- Probability, 102-40
 - Addition rules, 106
 - Binomial (see in B)
 - Empirical approach, 103
 - Formal mathematical approach, 102
 - Fundamental notions, 104
 - Multiplication rules, 105
 - Personalistic approach, 102
 - Possible outcomes, 104
- Product Moment Correlation
(see Correlational Techniques)
- Q_1 , 71
- Q_3 , 71, 72
- Quadrants, 21-22
- Quartile Deviation, Q , 70-76, 133
 - Def., 70
 - Calculation, 70-72
 - Properties, 72
- r , 142-46
- r^2 , 151
- R (see Multiple Correlation)
- Rank Difference Correlation, 147-51
- Range, 14
- Reaction Time, 130
- Regression, 154
- Regression and Prediction, 301-18
 - Assumptions, 311
 - History, 300-01
 - Multiple regression, 311-18
 - Raw score, 304-08
 - Simple regression, 304-11
- Reliability, 284-90
 - Alternate forms, 285-86
 - Rational-Equivalence, 287-88
 - Split-Half, 286-87
 - Test-Retest, 285
- ρ , 147-51
- Root Mean Square Deviation, 73
- Run Test, 225-27
- Sample, 79
- Scatter Plots, 142
- Scholastic Aptitude Test, 77
- Semi-Interquartile Range Q , 70-72
 - Cal., 70-72
 - Def., 70
 - Properties, 72
 - Uses, 72
- Sign Test, 218-22
 - Assumptions, 222
 - Large samples, 221-22
 - Uses, 222
- Significance of Difference
 - Between means, 183-92
 - Between proportions, 194-97
 - Between r 's, 197-99
 - Between standard deviations, 192-94
- Simultaneous Normal Equations, 304
- Skewness, 115, 131-32
 - Based on mean etc., 131
 - Based on percentiles, 132
 - Direction of, 131
 - Importance of, 136
 - Measure of, 131
 - Negative, 131
 - Positive, 131
 - Significance of, 136
- Sociability, 79
- Spearman Brown Prophecy Formula, 289
- Standard Deviation, 63-70
 - Calculation of, 63-70
 - Deviation score method, 65-66
 - Grouped scores, 66-69
 - Long method, 66-67
 - Short method, 67-69

- Raw score method, 65-66
- Standard error of, 178
- Ungrouped scores, 64-66
- Standard Error, 166-82
 - of Mean, 167-76
 - of Median, 176-78
 - of Percentage/Proportion, 178-79
 - of Standard deviation, 178
- Stanine Scale, 95-96
- Standard Scores, 77, 93-95
- Statistics
 - Def., 1-2
 - Descriptive, 3
 - Importance, 2-3
 - Inferential, 3
 - Sampling, 3
- t Distribution, 171-72
 - Basic formula, 171
 - Equation of, 171
- T Scale, T Scores, 96-100
 - Def., 96-97
 - Computation, 96-97
 - Merits, 97-100
- Tetrachoric Correlation, 159-61
- Total Absence Point, 79
- True Zero Point, 78
- Two-Tailed Tests, 199-200
- Type I Error, 200
- Type II Error, 200
- Unbiased Estimate of SD, 69
- Unimodal, 154
- Unit Normal Curve, 116
- Units of Uniform Size, 78, 79
- Validity, 290-94
 - Concurrent, 291-92
 - Content, 291
 - Construct, 293
 - Criterion related, 292-93
 - Face, 291
 - Factorial, 294
 - Factors affecting, 294
- Variables, 4
 - Continuous, 4
 - Def., 4
 - Discrete, 4
 - Dependent, 4
 - Independent, 4
- Variance, 65-70
 - Calculation, 65-69
 - Properties, 69-70
- Yates' Correction, 213-14
- Z, 197
- Z Conversion, 197
- Z Function, 180-81, 197





Dr Y.P. Aggarwal, a teacher with a brilliant academic career, obtained his Master's degree from Panjab University and M.Ed. degree from Kurukshetra University with distinction and a gold medal. He was awarded a Commonwealth Research Scholarship and subsequently a Ph.D. in Education at the University of Ottawa, Canada.

Dr Aggarwal has been teaching statistics and research methodology to post-graduate classes in Education since 1965. He has published four other books—*Sampling Methods for Social Investigation; Introduction to Statistics for Social Sciences (Ed.); Abstracts of M.Ed. and M.A. (Education) Dissertations (Ed.);* and *Educational Administration and Supervision (jointly)*. His forthcoming books are : *Research in Emerging Fields of Education (Ed.)* and *Better Sampling Methods*. In addition to contributing to a number of important reference books, he has published several research papers in Indian and American journals. He has been Professor of Education, Head and Dean at Kumaon University and Chairman, Department of Education, Kurukshetra University. At present he is Professor of Education at Kurukshetra University.